Applied Probability Trust (8 September 2017)

NOTE ON DISCOUNTED CONTINUOUS-TIME MARKOV DECISION PROCESSES WITH A LOWER BOUNDING FUNCTION

XIN GUO,* *University of Liverpool*

ALEXEY PIUNOVSKIY,* *University of Liverpool*

YI ZHANG,* *University of Liverpool*

Abstract

We consider the discounted continuous-time Markov decision process (CTMDP), where the negative part of each cost rate is bounded by a drift function, say w, whereas the positive part is allowed to be arbitrarily unbounded. Our focus is on the existence of a stationary optimal policy for the discounted CTMDP problems out of the more general class. Both constrained and unconstrained problems are considered. Our investigations are based on the continuous-time version of the Veinott transformation. This technique was not widely employed in the previous literature in CTMDPs, but it clarifies the roles of the imposed conditions in a rather transparent way.

Keywords: Continuous-time Markov decision processes; discounted criterion

2010 Mathematics Subject Classification: Primary 90C40

Secondary 60J25

1. Introduction

Discounted continuous-time Markov decision processes (CTMDPs) have been studied intensively since the 1960s, with one of the first works being [35]. Initially the

* Postal address: Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. E-mail: X.Guo21@liv.ac.uk.

* Postal address: Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. E-mail: piunov@liv.ac.uk.

* Postal address: Corresponding author. Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. E-mail: yi.zhang@liv.ac.uk.

theory is mainly developed for the finite state space models with bounded cost and transition rates. Later developments extend to models in a Borel state space with unbounded transition and cost rates, see e.g., [14, 20, 32]. When the cost rates are unbounded from both above and below, a standard setup is to assume that there is a weight (or Lyapunov) function say w, bounding the growth of the absolute value of the cost rates and the transition rates in a suitable sense, so that the value function will be also bounded by this function w. Then the investigation is based on the applicability of Dynkin's formula to the class of w-bounded functions, for which some additional conditions must be also imposed. This line of reasoning was followed and demonstrated in the recent monographs [21, 34] and the articles [4, 32]. If, as in the present paper, we only bound the growth of the negative part of each cost rate using the drift function w, which is thus called a lower bounding function, then the value function is in general not w-bounded. The approach based on the Dynkin's formula becomes less adequate.

On the other hand, now it is well known that a discounted CTMDP problem is equivalent to a total undiscounted DTMDP (discrete-time Markov decision process) problem with the same action space; see [13, 14]; see also [29, 22, 33] for the total undiscounted CTMDP problem. This approach has been applied to studying the discounted CTMDP problem with arbitrarily unbounded transition rate and nonnegative cost rates, see [14]. Nevertheless, the case, where the cost rates can take both positive and negative values, has never been treated with this approach, to the best of our knowledge. The reason is that when the transition rate is unbounded, the induced DTMDP is in general not absorbing in the sense of [1, 15], see Example 2 below. When the cost functions can take both positive and negative values, the studies of such DTMDPs, especially for constrained problems, are challenging, as demonstrated in [12], and are still underdeveloped, see e.g., [9].

Having said the above, discounted CTMDP problems with a lower bounding function have not been studied in the literature. The corresponding model in discounted discrete-time problems was treated in [3, 25, 26]. This type of cost functions appears in the literature of economics when one considers e.g., the logarithmic utility function, where it is put $-\ln(0) := \infty$, see Section 7 of [38]. Note that they can be reduced to equivalent discounted problems with nonnegative cost functions, see [39, 40], see also [1, 10]. We shall demonstrate the continuous-time version of this technique. In [3],

this type of model was studied for a specific piecewise deterministic Markov decision process with jumps driven by a Poisson process, but following a different method based on the Young topology, compared with the one here.

Our main contributions are as follows. Under conditions similar to those in [4], we show the existence of a deterministic stationary (respectively, stationary) optimal policy for the unconstrained (respectively, constrained) discounted CTMDP problems with a lower bounding function. Our argument is based on a transformation for non-homogeneous Markov pure jump processes, which, under some additional conditions, allows us to reduce the original problems to equivalent problems with nonnegative cost rates, so as for the reduction technique to apply. The roles of the additional conditions for this reduction are self-justified in a rather transparent way, as compared to the justification based on their relation to the Dynkin's formula, see [4], which considers only the discounted problem with a w-bounded cost rate in a denumerable state space, and is restricted to stationary policies. With the better understanding of the roles of the conditions, even in the specific case, where the cost rates are bounded by the drift function w, we improve the existing results in [20, 32] by withdrawing and weakening several conditions assumed therein.

The rest of the paper is organized as follows. In Section 2 we formulate the optimal control problems under consideration. The main statement is presented and proved in Section 3. Some auxiliary definitions and facts are included in the appendix.

2. Model description and problem statement

The objective of this section is to describe briefly the controlled process similarly to [13, 14, 27, 32], and the associated optimal control problem of interest in this paper.

In what follows, $\mathcal{B}(X)$ is the Borel σ-algebra of the Borel space X, I stands for the indicator function, and $\delta_{\{x\}}(\cdot)$ is the Dirac measure concentrated on the singleton $\{x\}$. A measure is σ-additive and $[0,\infty]$-valued. Below, unless stated otherwise, the term of measurability is always understood in the Borel sense. Throughout this article, we adopt the conventions of $\frac{0}{0} := 0,\ 0 \cdot \infty := 0,\ \frac{1}{0} := +\infty,\ \infty - \infty := \infty$.

The primitives of a CTMDP are the following elements $\{S, A, A(\cdot), q\}$, where S is a nonempty Borel state space, A is a nonempty Borel action space, the $\mathcal{B}(A)$-valued

multifunction $x \in S \to A(x)$ is, by assumption, with a measurable graph $\mathbb{K} := \{(x,a) \in S \times A : \ a \in A(x)\}$, and q stands for a signed kernel $q(dy|x,a)$ on $\mathcal{B}(S)$ given $(x,a) \in \mathbb{K}$ such that $\widetilde{q}(\Gamma|x,a) := q(\Gamma \setminus \{x\}|x,a) \geq 0$ for all $\Gamma \in \mathcal{B}(S)$. Throughout this paper, we assume that $q(\cdot|x,a)$ is conservative and stable, i.e., $q(S|x,a) = 0$, $\bar{q}_x = \sup_{a \in A(x)} q_x(a) < \infty$, where $q_x(a) := -q(\{x\}|x,a)$. The signed kernel q is often called the transition rate. Below we assume that the set $\mathbb{K}$ contains the graph of some measurable mapping from S to A.

Let us take the sample space Ω by adjoining to the countable product space $S \times ((0,\infty) \times S)^\infty$ the sequences of the form $(x_0, \theta_1, \dots, \theta_n, x_n, \infty, x_\infty, \infty, x_\infty, \dots)$, where $x_0, x_1, \dots, x_n$ belong to S, $\theta_1, \dots, \theta_n$ belong to $(0,\infty)$, and $x_\infty \notin S$ is the isolated point. We equip Ω with its Borel σ-algebra $\mathcal{F}$.

Let $t_0(\omega) := 0 =: \theta_0$, and for each $n \geq 0$, and each element $\omega := (x_0, \theta_1, x_1, \theta_2, \dots) \in \Omega$, let $t_n(\omega) := t_{n-1}(\omega) + \theta_n$, and $t_\infty(\omega) := \lim_{n\to\infty} t_n(\omega)$. Obviously, $t_n(\omega)$ are measurable mappings on $(\Omega, \mathcal{F})$. In what follows, we often omit the argument $\omega \in \Omega$ from the presentation for simplicity. Also, we regard x_n and θ_{n+1} as the coordinate variables, and note that the pairs $\{t_n, x_n\}$ form a marked point process with the internal history $\{\mathcal{F}_t\}_{t\geq 0}$, i.e., the filtration generated by $\{t_n, x_n\}$; see Chapter 4 of [27] for greater details. The marked point process $\{t_n, x_n\}$ defines the stochastic process on $(\Omega, \mathcal{F})$ of interest $\{\xi_t, t \geq 0\}$ by

$$\xi_t = \sum_{n\geq 0} I\{t_n \leq t < t_{n+1}\} x_n + I\{t_\infty \leq t\} x_\infty. \tag{1}$$

Here we accept $0 \cdot x := 0$ and $1 \cdot x := x$ for each $x \in S_\infty$, where $S_\infty := S \bigcup \{x_\infty\}$.

Definition 1. (a) A policy π for the CTMDP is a $\mathcal{P}(A)$-valued predictable process with respect to the internal history $\{\mathcal{F}_t\}$ so that

$$\begin{aligned} \pi(da|\omega,t) \ &= \ I\{t \geq t_\infty\}\delta_{a_\infty}(da) \\ &\quad + \sum_{n=0}^{\infty} I\{t_n < t \leq t_{n+1}\}\pi_n(da|x_0,\theta_1,\dots,\theta_n,x_n,t-t_n) \end{aligned}$$

for each $\omega = (x_0, \theta_1, x_1, \theta_2, \dots) \in \Omega$ and $t \in (0,\infty)$, where $a_\infty \notin A$ is some isolated point. Here, $\mathcal{P}(A)$ is the space of probability measures on $\mathcal{B}(A)$ endowed with the usual weak topology, and for each $n = 0, 1, 2, \dots$, $\pi_n(da|x_0, \theta_1, \dots, x_n, s)$

is a stochastic kernel on A concentrated on $A(x_n)$ given $x_0 \in S, \ldots, x_n \in S, s \in (0,\infty)$. We identify a policy π with the sequence of stochastic kernels $\{\pi_n\}_{n=0}^{\infty}$.

(b) A policy π is called Markov if, for some stochastic kernel φ on A concentrated on $A(x)$ from $(x,t) \in S \times (0,\infty)$, one can write $\pi(da|\omega,t) = \varphi(da|\xi_{t-},t)$ whenever $t < t_\infty$ A Markov policy is identified with the underlying stochastic kernel φ.

(c) A policy $\pi = \{\pi_n\}_{n=0}^{\infty}$ is called stationary if, with slight abuse of notations,

$$\pi_n(da|x_0,\theta_1,\ldots,x_n,s) = \pi(da|x_n)$$

for each of the stochastic kernels π_n. A stationary policy is further called deterministic if $\pi_n(da|x_0,\theta_1,\ldots,x_n,s) = \delta_{\{f(x_n)\}}(da)$ for some measurable mapping f from S to A such that $f(x) \in A(x)$ for each $x \in S$. We shall identify such a deterministic stationary policy with the underlying measurable mapping f.

The class of all policies for the CTMDP is denoted by Π, and the class of all Markov policies is Π^M.

Under a policy $\pi = \{\pi_n\}_{n=0}^{\infty} \in \Pi$, we define the following predictable random measure ν^π on $S \times (0,\infty)$ by

$$\begin{aligned} \nu^\pi(dt,dy) &:= \int_A \widetilde{q}(dy|\xi_{t-}(\omega),a)\pi(da|\omega,t)dt \\ &= \sum_{n\geq 0} \int_A \widetilde{q}(dy|x_n,a)\pi_n(da|x_0,\theta_1,\ldots,\theta_n,x_n,t-t_n)I\{t_n < t \leq t_{n+1}\}dt \end{aligned}$$

with $q_{x_\infty}(a_\infty) = q(dy|x_\infty,a_\infty) := 0 =: q_{x_\infty}(a)$ for each $a \in A$. Then, given the initial distribution γ, where γ is a probability measure on $\mathcal{B}(S)$, there exists a unique probability measure P_γ^π such that

$$P_\gamma^\pi(x_0 \in dx) = \gamma(dx),$$

and with respect to P_γ^π, ν^π is the dual predictable projection of the random measure associated with the marked point process $\{t_n, x_n\}$; see [24, 27]. Below, when γ is a Dirac measure concentrated at $x \in S$, we use the 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.

According to [24], the conditional distribution of (θ_{n+1}, x_{n+1}) with the condition on

$x_0, \theta_1, \ldots, \theta_n, x_n$ with $x_n \in S$ is given by

$$\begin{aligned}
& P_\gamma^\pi(\theta_{n+1} \in \Gamma_1,\ x_{n+1} \in \Gamma_2 | x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \\
= & \int_{\Gamma_1} e^{-\int_0^t \int_A q_{x_n}(a)\pi_n(da|x_0,\theta_1,\ldots,\theta_n,x_n,s)ds} \\
& \left\{ \int_A \widetilde{q}(\Gamma_2|x_n,a)\pi_n(da|x_0,\theta_1,\ldots,\theta_n,x_n,t) \right\} dt,\ \forall\ \Gamma_1 \in \mathcal{B}((0,\infty)),\ \Gamma_2 \in \mathcal{B}(S); \\
& P_\gamma^\pi(\theta_{n+1} = \infty,\ x_{n+1} = x_\infty | x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \\
= & e^{-\int_0^\infty \int_A q_{x_n}(a)\pi_n(da|x_0,\theta_1,\ldots,\theta_n,x_n,s)ds},
\end{aligned}$$

and for $x_n = x_\infty$ by

$$P_\gamma^\pi(\theta_{n+1} = \infty,\ x_{n+1} = x_\infty | x_0, \theta_1, x_1, \ldots, \theta_n, x_n) = 1.$$

Let $\infty > \alpha > 0$ be a fixed discount factor. For each $j = 0, 1, \ldots, N$, with $N \geq 1$ being a fixed integer, let c_j be a $(-\infty, \infty]$-valued measurable function on $\mathbb{K}$, representing a cost rate, and d_j be a fixed finite constant, representing a corresponding constraint. We shall consider the following unconstrained and constrained α-discounted optimal control problems for the CTMDP $\{S, A, A(\cdot), q\}$, respectively:

$$\text{Minimize over } \pi \in \Pi: \qquad E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_A c_0(\xi_t, a)\pi(da|\omega, t)dt \right],\ x \in S, \tag{2}$$

and

$$\begin{aligned}
\text{Minimize over } \pi \in \Pi: \qquad & E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_A c_0(\xi_t, a)\pi(da|\omega, t)dt \right] \\
\text{subject to} \qquad & E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_A c_j(\xi_t, a)\pi(da|\omega, t)dt \right] \leq d_j,\ j = 1, 2, \ldots, N.
\end{aligned} \tag{3}$$

Here and below, we put

$$c(x_\infty, a) := 0,\ \forall\ a \in A \bigcup \{a_\infty\}. \tag{4}$$

The conditions we impose below will ensure that the performance measures in the above two problems are well defined, though not necessarily finite.

A policy π^* is called optimal for the unconstrained problem (2) if

$$E_x^{\pi^*} \left[\int_0^\infty e^{-\alpha t} \int_A c_0(\xi_t, a)\pi^*(da|\omega, t)dt \right] = \inf_{\pi \in \Pi} E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_A c_0(\xi_t, a)\pi(da|\omega, t)dt \right]$$

for each $x \in S$. A policy π is called feasible for the constrained problem (3) if it satisfies all the inequalities therein. A feasible policy π for problem (3) is said to be of a finite value if

$$E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_A c_0^\pm(\xi_t, a)\pi(da|\omega, t)dt \right] < \infty.$$

A policy π^* is said to be optimal for problem (3) if it is feasible and satisfies

$$E_x^{\pi^*} \left[\int_0^\infty e^{-\alpha t} \int_A c_0(\xi_t, a)\pi^*(da|\omega, t)dt \right] \leq E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_A c_0(\xi_t, a)\pi(da|\omega, t)dt \right]$$

for each feasible policy π.

Note that the definition of optimality of a feasible policy for the constrained problem (3) requires a fixed initial state $x \in S$. Here, we did not consider the more general case of a fixed initial distribution just for brevity and readability. The case of a fixed initial distribution γ can be similarly treated with additional conditions regarding γ.

We would like to allow the possibility of cost rates unbounded from both above and below. We consider the following set of conditions to guarantee that the performance measures in problems (2) and (3) are well defined.

Condition 1. *There exists a $[1, \infty)$-valued measurable function w on S such that*

(a) for some finite constant $0 \leq \rho < \alpha$,

$$\int_S w(y)q(dy|x, a) \leq \rho w(x), \ \forall \ (x, a) \in \mathbb{K};$$

(b) for some finite constant $L > 0$,

$$c_i^-(x, a) \leq Lw(x), \ \forall \ (x, a) \in \mathbb{K}, \ i = 0, 1, \ldots, N.$$

Here, for each $i = 0, 1, \ldots, N$, c_i^- is the negative part of the function c_i.

Below, we allow that $w(x_\infty) := 0$. The cost rates satisfying part (b) of the above condition are said to be lower bounded by the drift function w; c.f. p.251 of [3] for a related definition for piecewise deterministic Markov decision processes.

Lemma 1. *Suppose Condition 1 is satisfied. Let a policy π be arbitrarily fixed. Then*

$$E_x^\pi \left[\int_0^\infty e^{-\alpha t} w(\xi_t)dt \right] < \infty, \ \forall \ x \in S.$$

In particular, for each $x \in S$, the integrals $E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_A c_i(\xi_t, a)\pi(da|\omega, t)dt \right]$, $i = 0, 1, \ldots, N$, are well defined.

Proof. This follows from Lemma 2 of [31] and (4). □

Assumption 1. *Throughout this paper, unless stated otherwise, Condition 1 is assumed to hold automatically, without specific reference.*

3. Main statement and its proof

3.1. Conditions, statements and comments

Condition 2. *There exist a $(0,\infty)$-valued measurable function w' on S and a monotone nondecreasing sequence of measurable subsets $\{Z_m\}_{m=1}^{\infty} \subseteq \mathcal{B}(S)$ such that the following hold.*

(a) $Z_m \uparrow S$ *as* $m \to \infty$.

(b) $\sup_{x\in Z_m} \overline{q}_x < \infty$ *for each* $m = 1, 2, \ldots$.

(c) For some constant $\rho' \in (0,\infty)$,

$$\int_S w'(y) q(dy|x,a) \le \rho' w'(x), \ \forall\ x \in S,\ a \in A(x).$$

(d) $\inf_{x\in S\setminus Z_m} \frac{w'(x)}{w(x)} \to \infty$ *as* $m \to \infty$, *where the function* w *comes from Condition 1.*

Let a $[0,\infty)$-valued function v on S be fixed. A function g on S is called v-bounded if $||g||_v := \sup_{x\in S} \frac{|g(x)|}{v(x)} < \infty$; here the convention of $0/0 = 0$ is in use.

Condition 3. *(a) The multifunction $x \in S \to A(x) \in \mathcal{B}(A)$ is compact-valued and upper semicontinuous.*

(b) For each w-bounded continuous function g on S, $(x,a) \in \mathbb{K} \to \int_S g(y)\tilde{q}(dy|x,a)$ is continuous. Here and below the function w is from Condition 1.

(c) The function w is continuous on S, and the functions c_i are lower semicontinuous on $\mathbb{K}$.

The conditions formulated in the above can be satisfied when the negative part of each cost rate is bounded by a drift function, whereas the positive part is arbitrarily unbounded. In the literature of economics, such a cost rate might appear e.g., when one considers the logarithmic utility function, where they put $-\ln 0 := \infty$, see Section

7 of [38]; see also Example 2 of [26]. We formulate an example of such a CTMDP as follows.

Example 1. Consider a controlled M/M/∞ queueing system. We put $S = \{0, 1, \dots\}$. The state $x \in S$ represents the number of customers in the system. The control is the arrival rate $a \in [0, x] \subseteq [0, \infty)$ for each $x \in S$. The service rate $\mu > 0$ is uncontrolled. The cost rate is given by $c_0(x, a) = -\ln a$, and the constraint cost rate is given by $c_1(x, a) = x$. Then Conditions 1, 2 and 3 are all satisfied (for a large enough discount factor); one can put $w(x) = x + 1$ and $w'(x) = 1 + x^2$. On the other hand, there is no finite bounding function for $|c_0|$.

The next condition is for constrained problem only.

Condition 4. *There exists a feasible policy for problem (3) with a finite value.*

The main statement of this paper is the following one.

Theorem 1. *Suppose Conditions 1, 2 and 3 are satisfied. Then the following assertions hold.*

(a) There exists a deterministic stationary optimal policy for the unconstrained problem (2). In fact, one can always take a deterministic stationary policy providing the minimum in the equation (15) as a deterministic stationary optimal policy.

(b) If Condition 4 is also satisfied, then there exists a stationary optimal policy for the constrained problem (3).

In the previous literature, general discounted CTMDPs have not been considered when the cost rates were bounded below by a lower bounding function, and arbitrarily unbounded from the above, although for specific piecewise deterministic Markov decision processes with jumps driven by a Poisson process, this was considered in [3] following a different method. Discrete-time problems with a lower bounding function were considered in [3, 25], and in latter reference, the motivation for considering such cost functions was explained with their applications to economics. For discounted DTMDP problems, the treatment in [3, 25] was direct. But it is possible to reduce this to equivalent problems with nonnegative cost functions, using the technique in

p.101 of [40], see also [10] and p.79 of [1]. The proof of Theorem 1 will be based on a similar technique for CTMDPs, which, to the best of our knowledge, has not been widely applied to CTMDPs.

For the more restrictive case, where the cost rates are w-bounded, with w coming from Condition 1, Theorem 1(a) was obtained in [4] under essentially equivalent conditions for discounted CTMDPs in a denumerable state space but restricted to the class of stationary policies. Here we show that it is without loss of generality to be restricted to this narrower class of policies under the imposed conditions. Otherwise, this sufficiency result seems not to follow from other known results in the relevant literature. The approach in [4] was directly based on the application of the Dynkin's forumla, and is different from ours. When the cost rates are only lower w-bounded, the value function is in general not w-bounded. Since under the conditions in [4] and here, Dynkin's formula is only applicable to the class of w-bounded functions, the treatment in [4] does not directly apply to the general case dealt with here.

Also when the cost rates are w-bounded, Theorem 1(b) was obtained in e.g., [32] but under stronger conditions. We include them here for ease of reference.

Instead of Condition 2, the following condition was imposed in [32].

Condition 5. *There exists a $(0,\infty)$-valued measurable function $\widetilde{w}'$ on S such that the following hold.*

(a) For some constant $\widetilde{L}' \in (0,\infty)$, $\overline{q}_x \le \widetilde{L}'\widetilde{w}'(x)$ for each $x \in S$.

(b) For some constant $\widetilde{\rho}' \in (0,\infty)$, $\int_S \widetilde{w}'(y)q(dy|x,a) \le \widetilde{\rho}'\widetilde{w}'(x)$ for each $(x,a) \in \mathbb{K}$.

(c) For some constant $\widetilde{L} \in (0,\infty)$, $(\overline{q}_x + 1)w(x) \le \widetilde{L}\widetilde{w}'(x)$ for each $x \in S$, where the function w comes from Condition 1.

It is easy to see that, if the above condition is satisfied, then so is Condition 2 with $w' = \widetilde{w}' + 1$, $\rho' = \widetilde{\rho}'$, $Z_m = \left\{x \in S : \ \frac{\widetilde{w}'(x)+1}{w(x)} \le m\right\}$ for each $m = 1, 2, \ldots$.

Furthermore, under Conditions 1, 2 and 4, in addition to Condition 3, it was also assumed in [32] that the function $\frac{\widetilde{w}'}{w}$ is a moment function on $\mathbb{K}$, see Definition E.7 of [23], in order to apply the Prokhorov theorem in their proof, see Proposition E.8 and Theorem E.6 of [23]. This is not needed here. The investigations in [32] are

largely based on the Dynkin's formula, and do not handle the more general cost rates considered here.

The rest of this section proves Theorem 1. On the way, we comment and clarify the roles of the imposed conditions, and present the auxiliary statements.

3.2. Proof of the main statement

The proof of Theorem 1 follows from a sequence of lemmas. The outline of the proof steps is announced in the next remark.

Remark 1. The main themes in the proof of Theorem 1 can be summarized as follows.

1. Under Condition 1, the w-transformation, see Lemma 3, allows one to reduce the original problems (2) and (3) to problems (6) and (7) for the w-transformed CTMDP model with cost rates bounded from below, equivalently.

2. Under the extra Condition 2, problems (6) and (7) are reduced to discounted CTMDP problems (9) and (10) with nonnegative cost rates by adding some large enough constant. This is possible because Condition 2 ensures that the controlled process in the w-transformed CTMDP model is nonexplosive under each Markov policy, according to Lemma 4.

3. By applying the reduction technique in [13, 14], discounted CTMDP problems (9) and (10) with nonnegative cost rates are reduced to total undiscounted DTMDP problems (13) and (14) with nonnegative cost functions.

4. Apply the optimality results in [8] to the DTMDP problems (13) and (14) with nonnegative cost functions. Then deduce from here the corresponding optimality results for the original problems (2) and (3).

The details are as follows.

Proof of Theorem 1. The following statement is a consequence of Theorem 4.2 of [16], see also [18], and is the starting point of our reasoning.

Lemma 2. *For each initial state $x \in S$ and policy π, there exists a Markov policy φ such that*

$$E_x^{\pi}\left[\int_0^{\infty} e^{-\alpha t}\int_A f(\xi_t,a)\pi(da|\omega,t)dt\right] = E_x^{\varphi}\left[\int_0^{\infty} e^{-\alpha t}\int_A f(\xi_t,a)\varphi(da|\xi_t,t)dt\right]$$

for each $[0,\infty]$*-valued measurable function* f *on* $\mathbb{K}$.

The above lemma implies that without loss of generality, we can restrict to the class of Markov policies for problems (2) and (3), i.e., if we obtain an optimal policy out of the class of Markov policies for problem (2) (or (3)), then that policy is optimal for problem (2) (or (3)) out of the general class.

We recall some definitions related to the process $\{\xi_t, t \geq 0\}$ under a Markov policy φ. Let us consider the signed kernel on S from $S \times [0,\infty)$ defined by

$$q_\varphi(dy|x,t) := \int_A q(dy|x,a)\varphi(da|x,t),\ \forall\ x \in S,\ t \in [0,\infty).$$

Then q_φ is a conservative and stable Q-function in the sense of [17, p.262]. For the ease of reference, we recall some relevant definitions and facts about Q-functions in the appendix.

According to Theorem 2.2 of [17], under a Markov policy, say φ, the process $\{\xi_t, t \geq 0\}$ is a Markov pure jump process on $\{\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P^\varphi\}$, that is, for each $s,t \in [0,\infty)$,

$$P^\varphi(\xi_{t+s} \in \Gamma|\mathcal{F}_t) = P^\varphi(\xi_{t+s} \in \Gamma|\xi_t),\ \forall\ \Gamma \in \mathcal{B}(X_\infty);$$

and each trajectory of $\{\xi_t; t \geq 0\}$ is piecewise constant and right-continuous, such that for each $t \in [0, t_\infty)$, there are finitely many discontinuity points on the interval $[0,t]$, see Definition 1 in Chapter III of [19]. Here and below, we omit the subscript in P^φ_γ, whenever the initial distribution γ is irrelevant. Furthermore, by Theorem 2.2 of [17], p_{q_φ} defined by (17) with q being replaced by q_φ is the transition function corresponding to the process $\{\xi_t, t \geq 0\}$, i.e., for each $s \leq t$, on $\{s < t_\infty\}$,

$$P^\varphi(\xi_t \in \Gamma|\mathcal{F}_s) = p_{q_\varphi}(s,\xi_s,t,\Gamma),\ \forall\ \Gamma \in \mathcal{B}(S),$$

c.f. p.1397 of [28]. Consequently, for each Markov policy φ,

$$\begin{aligned} & E^\varphi_x\left[\int_0^\infty e^{-\alpha t}\int_A c_i(\xi_t,a)\varphi(da|\xi_t,t)dt\right] \\ = & \int_0^\infty\int_S e^{-\alpha t}\int_A c_i(y,a)\varphi(da|y,t)p_{q_\varphi}(0,x,t,dy)dt,\ \forall\ x \in S \end{aligned}$$

for each $i = 0,1,\ldots,N$.

Given the Q-function q_φ on S induced by a Markov policy φ, let us introduce the w-transformed Q-function q^w_φ on S_δ defined as follows.

Let

$$S_\delta := S \bigcup \{\delta\}$$

with $\delta \notin S$ being an isolated point concerning the topology of S_δ that satisfies $\delta \neq x_\infty$. The w-transformed (stable conservative) Q-function q^w_φ on S_δ is defined by

$$q^w_\varphi(\Gamma|x,s) := \begin{cases} \frac{\int_\Gamma w(y) q_\varphi(dy|x,s)}{w(x)}, & \text{if } x \in S,\ \Gamma \in \mathcal{B}(S),\ x \notin \Gamma; \\ \rho - \frac{\int_S w(y) q_\varphi(dy|x,s)}{w(x)}, & \text{if } x \in S,\ \Gamma = \{\delta\}; \\ 0, & \text{if } x = \delta,\ \Gamma = S_\delta. \end{cases} \tag{5}$$

for each $s \in [0,\infty)$; and

$$q^w_{\varphi\, x}(s) := \rho + q_{\varphi\, x}(s),\ \forall\ s \in [0,\infty).$$

Here, $q_{\varphi\, x}(s) = -q_\varphi(S \setminus \{x\}|x,s)$; see the appendix for more definitions and relevant notations concerning a Q-function. This transformation is the continuous-time version of the Veinott transformation, see [39], widely known in the literature of DTMDPs. For (uncontrolled) homogeneous continuous-time Markov chains, this transformation was used in e.g., [2, 36, 37].

Lemma 3. *Let a Markov policy φ be fixed. For each $x \in S$, $s,t \in [0,\infty)$, $s \leq t$ and $\Gamma \in \mathcal{B}(S)$, the following relation holds;*

$$p_{q^w_\varphi}(s,x,t,\Gamma) = \frac{e^{-\rho(t-s)}}{w(x)} \int_\Gamma w(y) p_{q_\varphi}(s,x,t,dy).$$

Proof. See Lemma A.3 of [41]. □

By Lemma 3, we see that for each $i = 0,1,\ldots,N$,

$$\begin{aligned} & w(x) \int_0^\infty \int_S p_{q^w_\varphi}(0,x,t,dy) \int_A \frac{c_i(y,a)}{w(y)} \varphi(da|y,t) e^{-(\alpha-\rho)t} dt \\ = & \int_0^\infty \int_S \int_A c_i(y,a) \varphi(da|y,t) e^{-\alpha t} p_{q_\varphi}(0,x,t,dy) dt,\ \forall\ x \in S. \end{aligned}$$

Hence, problem (2) is equivalent to

$$\text{Minimize over } \varphi \in \Pi^M\text{:} \quad \int_0^\infty \int_S p_{q^w_\varphi}(0,x,t,dy) \int_A \frac{c_0(y,a)}{w(y)} \varphi(da|y,t) e^{-(\alpha-\rho)t} dt, \quad x \in S, \tag{6}$$

and problem (3) is equivalent to

$$\begin{aligned} \text{Minimize over } \varphi\in\Pi^M: \quad & \int_0^\infty\int_S p_{q^w_\varphi}(0,x,t,dy)\int_A \frac{c_i(y,a)}{w(y)}\varphi(da|y,t)e^{-(\alpha-\rho)t}dt \\ \text{subject to} \quad & \int_0^\infty\int_S p_{q^w_\varphi}(0,x,t,dy)\int_A \frac{c_j(y,a)}{w(y)}\varphi(da|y,t)e^{-(\alpha-\rho)t}dt \\ & \le \frac{d_j}{w(x)},\ j=1,2,\dots,N. \end{aligned} \tag{7}$$

Thus, one can consider the w-transformed CTMDP $\{S_\delta, A\bigcup\{a_\infty\}, A_\delta(\cdot), q^w\}$, where $A_\delta(\delta) := \{a_\infty\}$, and $A_\delta(x) := A(x)$ for each $x \in S$, while the transition rate q^w is defined by, c.f. (5),

$$q^w(\Gamma|x,a) = \begin{cases} \frac{\int_\Gamma w(y)q(dy|x,a)}{w(x)}, & \text{if } x\in S,\ \Gamma\in\mathcal{B}(S),\ x\notin\Gamma; \\ \rho - \frac{\int_S w(y)q(dy|x,a)}{w(x)}, & \text{if } x\in S,\ \Gamma=\{\delta\}; \\ 0, & \text{if } x=\delta,\ \Gamma=S_\delta. \end{cases}$$

for each $x \in S_\delta$ and $a \in A_\delta(x)$; and

$$q^w_x(a) := \rho + q_x(a),\ \forall\ x\in S,\ a\in A_\delta(x).$$

The requirement of $\alpha > \rho$ in Condition 1(a) is needed so that problems (6) and (7) are legitimate $(\alpha-\rho)$-discounted problems of the w-transformed CTMDP with the cost rates c^w_i defined by

$$c^w_i(x,a) := \frac{c_i(x,a)}{w(x)}$$

for each $x \in S$, $a \in A(x)$; and

$$c^w_i(\delta,a_\infty) := 0.$$

According to the reduction technique for discounted CTMDPs, see [14], the CTMDP problems (6) and (7) can be reduced to equivalent total undiscounted problems for the DTMDP $\{S_\delta\bigcup\{x_\infty\}, A\bigcup\{a_\infty\}, A_\delta(\cdot), T\}$ with the cost functions C_i, where the transition probability T is defined by

$$T(\Gamma|x,a) := \frac{\int_\Gamma w(y)q(dy|x,a)}{(\alpha+q_x(a))w(x)}$$

for each $\Gamma\in\mathcal{B}(S)$, $x\notin\Gamma$, and $a\in A_\delta(x)$;

$$T(\{\delta\}|x,a) := \frac{\rho w(x) - \int_S w(y)q(dy|x,a)}{(\alpha+q_x(a))w(x)}$$

for each $x \in S$ and $a \in A_\delta(x)$;

$$T(\{x_\infty\}|x,a) := \frac{\alpha-\rho}{\alpha+q_x(a)}$$

for each $x \in S$ and $a \in A_\delta(x)$; and $T(\{x_\infty\}|x_\infty,a_\infty) := 1 =: T(\{x_\infty\}|\delta,a_\infty)$, and the cost functions C_i are defined by

$$C_i(x,a) := \frac{c_i(x,a)}{(\alpha+q_x(a))w(x)}$$

for each $x \in S$ and $a \in A_\delta(x)$; and

$$C_i(\delta,a_\infty) := 0 =: C_i(x_\infty,a_\infty).$$

More precisely, given the initial state $x \in S$, for each Markov policy φ for the w-transformed CTMDP, there is a strategy σ for the DTMDP $\{S_\delta \bigcup \{x_\infty\}, A \bigcup \{a_\infty\}, A_\delta(\cdot), T\}$ such that

$$\int_0^\infty \int_S p_{q_\varphi^w}(0,x,t,dy)\frac{c_i(y,a)}{w(y)}e^{-(\alpha-\rho)t}dt = \mathbb{E}_x^\sigma\left[\sum_{n=0}^\infty C_i(X_n,A_n)\right]$$

for each $i = 0,1,\dots,N$, and vice versa. Moreover, in the previous equality, if φ is a deterministic stationary (respectively, stationary) policy, then σ can be taken as a deterministic stationary (respectively, stationary) strategy for the DTMDP, and vice versa. Here we use $\mathbb{E}_x^\sigma$ to denote the expectation taken with respect to the strategic measure of the DTMDP under the strategy σ, and $\{X_n\}$ and $\{A_n\}$ are the controlled and controlling processes in the DTMDP. The term "strategy" is reserved for the DTMDP to avoid the potential confusion with the corresponding notion for the CTMDP. We refer the reader to e.g., [23, 30] for the standard description of a DTMDP.

Note that in general, the DTMDP $\{S_\delta \bigcup \{x_\infty\}, A \bigcup \{a_\infty\}, A_\delta(\cdot), T\}$ is not absorbing in the sense of [1, 15], and the cost function C_i can take both positive and negative values. We formulate such a CTMDP in the next example.

Example 2. Suppose the CTMDP is an uncontrolled pure birth process with $S = \{1,2,\dots\}$. The birth rate at the state $x \in S$ is $2x$. The discount factor is $\alpha = 2$. We put $\rho = 0$ and $w(x) = 1$ for each $x \in S$. Suppose the cost rate is only zero at the state δ. For the induced DTMDP, $\{x_\infty\}$ is the absorbing set; the point δ can be excluded from the state space because it is never reached starting from $S \bigcup \{x_\infty\}$. Then one can

show that starting from 1, the expected time until the DTMDP reaches x_∞ is infinite. In accordance with e.g., [1, 15], this means that the model is not absorbing, i.e., the expected time to absorption is not finite.

On the other hand, the functions c_i^w, $i = 0, 1, \dots, N$, are bounded from below under Condition 1(b). Let some common lower bound be $\underline{c} \le 0$. Let

$$\widetilde{c}_i^w := c_i^w - \underline{c} \tag{8}$$

for each $i = 0, 1, \dots, N$. Then the functions $\widetilde{c}_i^w$ are all nonnegative. In order for problems (6) and (7) to be equivalent to

$$\text{Minimize over } \varphi \in \Pi^M\text{:} \qquad \int_0^\infty \int_{S_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{A_\delta} \widetilde{c}_0^w(y, a)\varphi(da|y, t)e^{-(\alpha-\rho)t}dt, \quad x \in S, \tag{9}$$

and

$$\begin{aligned} \text{Minimize over } \varphi \in \Pi^M\text{:} \qquad & \int_0^\infty \int_{S_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{A_\delta} \widetilde{c}_0^w(y)\varphi(da|y, t)e^{-(\alpha-\rho)t}dt \\ \text{such that} \qquad & \int_0^\infty \int_{S_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{A_\delta} \widetilde{c}_j^w(y)\varphi(da|y, t)e^{-(\alpha-\rho)t}dt \\ & \le \frac{d_j}{w(x)} - \frac{\underline{c}}{\alpha - \rho}, \ j = 1, 2, \dots, N, \end{aligned} \tag{10}$$

respectively, we need the following relation to hold for each $\varphi \in \Pi^M$:

$$p_{q_\varphi^w}(0, x, t, S_\delta) = 1, \ \forall\, x \in S, \ t \in [0, \infty). \tag{11}$$

In general, problems (6) and (7) are not equivalent to problems (9) and (10). We demonstrate this with the following example, which was also considered by Spieksma in [37].

Example 3. Let $S = \{0, 1, 2, \dots\}$, $A(x) \equiv A = \{0, 1\}$. We endow them with the discrete topology. The transition rate is given by

$$q(\{y\}|x, 0) = \begin{cases} \frac{5}{12}2^x, & \text{if } x \neq 0, \ y = x + 1; \\ \frac{7}{12}2^x, & \text{if } x \neq 0, \ y = x - 1; \\ 0, & \text{if } x = 0. \end{cases}$$

and $q(\{y\}|x,1)=0$ for each $x,y\in S$. Let $w(x)=\left(\frac{7}{5}\right)^x$ for each $x\in S$. Then one can verify that

$$\sum_{y\in S} w(y)q(\{y\}|x,a)=0,\ \forall\ x\in S,\ a\in A,$$

and so let $\rho=0$, and $\alpha=1$. Let $c_0(x,a)\equiv 0$. Put $\underline{c}=-1$. Conditions 1 and 3 are satisfied.

Now

$$q^w(\{y\}|x,0)=\begin{cases}\frac{7}{12}2^x, & \text{if } x\neq\delta,\ x\neq 0,\ y=x+1;\\ \frac{5}{12}2^x, & \text{if } x\neq\delta,\ x\neq 0,\ y=x-1;\\ 0, & \text{if } x\neq\delta,\ y=\delta;\\ 0, & \text{if } x=\delta \text{ or } x=0.\end{cases}$$

and $q^w_x(0)=2^x$ for each $x\neq\delta,\ 0$, and $q^w_x(0)=0$ if $x=0,\ \delta$. Also $q^w_x(1)=0$ for each $x\in S_\delta$.

Consider the following two deterministic stationary strategies: $\varphi_0(da|x,t)\equiv\delta_0(da)$ and $\varphi_1(da|x,t)\equiv\delta_1(da)$. Clearly, they are both optimal for problem (6). On the other hand,

$$\begin{aligned}&\int_0^\infty\int_{S_\delta} p_{q^w_{\varphi_i}}(0,x,t,dy)\int_{A_\delta}\widetilde{c}^w_0(y,a)\varphi_i(da|y,t)e^{-(\alpha-\rho)t}dt\\ =\ &\int_0^\infty p_{q^w_{\varphi_i}}(0,x,t,S_\delta)e^{-t}dt,\ x\in S,\ i=0,1.\end{aligned}$$

Clearly, $p_{q^w_{\varphi_1}}(0,x,t,S_\delta)\equiv 1=\int_0^\infty p_{q^w_{\varphi_1}}(0,x,t,S_\delta)e^{-t}dt$. It is shown in Section 5 of [37] that (11) does not hold for $\varphi=\varphi_0$ with some $x\in S$; this can also be checked using Theorem 2 of [5]. It follows that for some $x\in S$, $\int_0^\infty p_{q^w_{\varphi_0}}(0,x,t,S_\delta)e^{-t}dt<1$; see also Lemma 2.1 of [41]. Therefore, the policy φ_1 is not optimal for problem (9), although it is optimal for problem (6). Hence, in general, (6) and (7) are not equivalent to problems (9) and (10).

Remark 2. Example 3 illustrates the role of the requirement (11). Condition 2 is precisely imposed for this purpose, as seen in the next statement. (An alternative justification of the role of Condition 2 is that it validates the Dynkin's formula for the original CTMDP to a certain class of functions, see [4] for the homogeneous denumerable case. But the explanation here is more transparent in our opinion.) In

the literature, e.g., [20, 32, 34], stronger conditions, e.g., Condition 5, than Condition 2, were imposed to guarantee (11) to hold. The investigations there were not based on reduction method to DTMDP.

Lemma 4. *Let some Markov policy φ be fixed. Suppose Condition 1(a) and Condition 2 are satisfied. Then (11) holds.*

Proof. According to Theorem 2, for the statement it suffices to verify that Condition 6 is satisfied.

Since the Markov policy φ is fixed throughout this proof, we write q_φ as q for brevity. Note that

$$\begin{aligned}&\int_S \frac{w'(y)}{w(y)} q^w(dy|x,s) = \int_S \frac{w'(y)}{w(y)} \frac{w(y)}{w(x)} \widetilde{q}(dy|x,s) - (\rho + q_x(s)) \frac{w'(x)}{w(x)} \\ = &\int_S \frac{w'(y)}{w(y)} \widetilde{q}(dy|x,s) - (\rho + q_x(s)) \frac{w'(x)}{w(x)} \le (\rho' - \rho) \frac{w'(x)}{w(x)}, \ \forall\ x \in S,\ s \ge 0. \quad (12)\end{aligned}$$

Consider the $[0,\infty)$-valued measurable function $\widetilde{w}$ on $[0,\infty) \times S_\delta$ defined for each $v \in [0,\infty)$ by $\widetilde{w}(v,x) = \frac{w'(x)}{w(x)}$ if $x \in S$ and $\widetilde{w}(v,\delta) = 0$. Then Condition 6, with S and q being replaced by S_δ and q^w, is satisfied by the monotone nondecreasing sequence of measurable subsets $\{\widetilde{V}_n\}_{n=1}^\infty$ of $\mathbb{R}_+^0 \times S_\delta$ defined by $\widetilde{V}_n = [0,\infty) \times V_n \bigcup \{\delta\}$ for each $n = 1, 2, \ldots$, and the function $\widetilde{w}$ on $[0,\infty) \times S_\delta$ defined in the above. In greater detail, part (d) of the corresponding version of Condition 6 is satisfied because, by (12),

$$\begin{aligned}&\int_0^\infty \int_{S_\delta} \widetilde{w}(t+v,y) e^{-\rho' t - \int_{(0,t]} q_x^w(s+v)ds} \widetilde{q^w}(dy|x,t+v)dt \\ \le &\int_0^\infty e^{-\rho' t - \int_0^t q_x^w(s+v)ds} \left(q_x(s) + \rho'\right) \widetilde{w}(v,x) = \widetilde{w}(v,x),\ \forall\ x \in S,\end{aligned}$$

and the last inequality holds trivially when $x = \delta$.

Thus, by Theorem 2, we see that relation (11) is satisfied, and the statement follows. □

By the way, under Condition 1(a), in certain models, Condition 2 is also necessary for (11) to hold under certain policies; see [41]. In the homogeneous denumerable case, this was first observed in [36]. For more concrete examples such as single birth processes, this necessity part was known earlier, see [6].

As a result of the above lemma and the discussions above it, we see that under Condition 1 and Condition 2, one can reduce the α-discounted problems (2) and (3) for

the original CTMDP $\{S, A, A(\cdot), q\}$ to the $(\alpha-\rho)$-discounted problems (9) and (10) for the CTMDP $\{S_\delta, A_\delta, A_\delta(\cdot), q^w\}$ with nonnegative cost rates. Furthermore, according to the reduction technique [14], which was also sketched in the above, problems (9) and (10) can be reduced to

$$\text{Minimize over } \sigma \qquad \mathbb{E}_x^\sigma \left[\sum_{n=0}^{\infty} \widetilde{C}_0(X_n, A_n)\right], \ x \in S, \tag{13}$$

and

$$\begin{aligned} \text{Minimize over } \sigma\text{:} \qquad & \mathbb{E}_x^\sigma \left[\sum_{n=0}^{\infty} \widetilde{C}_0(X_n, A_n)\right] \\ \text{such that} \qquad & \mathbb{E}_x^\sigma \left[\sum_{n=0}^{\infty} \widetilde{C}_j(X_n, A_n)\right] \le \frac{d_j}{w(x)} - \frac{c}{\alpha-\rho}, \\ & j = 1, 2, \ldots, N, \end{aligned} \tag{14}$$

respectively, for the DTMDP $\{S_\delta \bigcup \{x_\infty\}, A \bigcup \{a_\infty\}, A_\delta(\cdot), T\}$ defined earlier. Here the cost functions $\widetilde{C}_i$ for the DTMDP are defined by

$$\widetilde{C}_i(x, a) := \frac{\widetilde{c}_i^w(x, a)}{(\alpha + q_x(a))} \ge 0$$

for each $x \in S_\delta$ and $a \in A_\delta(x)$; and

$$\widetilde{C}_i(x_\infty, a_\infty) := 0,$$

with the functions $\widetilde{c}_i^w$ being defined by (8). Note that the cost functions $\widetilde{C}_i$ could be arbitrarily unbounded from above.

Finally, if Condition 1, Condition 2, and Condition 3 are all satisfied, then it is easy to check that the DTMDP $\{S_\delta \bigcup \{x_\infty\}, A \bigcup \{a_\infty\}, A_\delta(\cdot), T\}$ with the nonnegative cost functions $\widetilde{C}_i$ is a semicontinuous model, see [3, 11], and it is a standard result that there exists an optimal deterministic stationary strategy for problem (13). For the constrained problem (14), under the extra Condition 4, one can refer to Theorem 4.1 of [8], see also Theorem A.2 of [7], for the existence of a stationary optimal strategy for (14). Since these two DTMDP problems are equivalent to the original CTMDP problems, according to the reduction technique for discounted CTMDP problems as mentioned earlier, we immediately conclude the existence of an optimal deterministic stationary policy for the unconstrained CTMDP problem (2) and an optimal stationary

policy for the constrained CTMDP problem (3). The proof of Theorem 1 is thus completed. □

We finish this section with the following observation. Suppose Conditions 1 and 3 are satisfied. If one solves problem (9) with a deterministic stationary policy φ, which also satisfies (11), then φ is also optimal for problem (6), in spite that Condition 2 has not been assumed to hold uniformly in all actions.

The justifications of this claim are as follows. In general, problems (6) and (7) are not equivalent to (9) and (10), respectively; recall Example 3. According to [14], (9) is equivalent to the DTMDP problem $\{S_\delta \bigcup \{x_\infty\}, A \bigcup \{a_\infty\}, A_\delta(\cdot), T\}$ with the cost function $\widetilde{C}_0$. Suppose φ^* is an optimal deterministic strategy for this DTMDP problem. Under Conditions 1 and Condition 3, if V^* denotes the value function of this DTMDP problem, then such an optimal deterministic stationary strategy exists and can be obtained by taking the measurable selector providing the minimum in the following:

$$V^*(x) = \inf_{a \in A_\delta(x)} \left\{ \widetilde{C}_0(x,a) + \int_{S_\delta} T(dy|x,a) V^*(y) \right\}, \ \forall \ x \in S_\delta. \tag{15}$$

We claim that φ^* is also an optimal deterministic policy for the CTMDP problem (6), provided that (11) holds for this particular strategy φ^*, i.e.,

$$p_{q^w_{\varphi^*}}(0,x,t,S_\delta) = 1, \ \forall \ x \in S, \ t \in [0,\infty). \tag{16}$$

Indeed, since φ^* is optimal for the DTMDP $\{S_\delta \bigcup \{x_\infty\}, A \bigcup \{a_\infty\}, A_\delta(\cdot), T\}$ with the cost function $\widetilde{C}_0$, which is equivalent to problem (9),

$$\begin{aligned}
& \inf_{\varphi \in \Pi^M} \left\{ \int_0^\infty \int_{S_\delta} p_{q^w_\varphi}(0,x,t,dy) \int_{A_\delta} \widetilde{c}^w_0(y,a) \varphi(da|y,t) e^{-(\alpha-\rho)t} dt \right\} \\
= & \int_0^\infty \int_{S_\delta} p_{q^w_{\varphi^*}}(0,x,t,dy) \widetilde{c}^w_0(y,\varphi^*(y)) e^{-(\alpha-\rho)t} dt \\
= & \int_0^\infty \int_{S} p_{q^w_{\varphi^*}}(0,x,t,dy) \frac{c_0(y,\varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt - \frac{\underline{c}}{\alpha-\rho}, \ \forall \ x \in S.
\end{aligned}$$

Consider an arbitrarily fixed $\varphi \in \Pi^M$. Then for each $x \in S$,

$$\begin{aligned}
& \int_0^\infty \int_S p_{q^w_{\varphi^*}}(0,x,t,dy) \frac{c_0(y,\varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt - \frac{\underline{c}}{\alpha-\rho} \\
\leq & \int_0^\infty \int_{S_\delta} p_{q^w_{\varphi}}(0,x,t,dy) \int_{A_\delta} \widetilde{c}^w_0(y,a) \varphi(da|y,t) e^{-(\alpha-\rho)t} dt \\
= & \int_0^\infty \int_S p_{q^w_{\varphi}}(0,x,t,dy) \int_A \frac{c_0(y,a)}{w(y)} \varphi(da|y,t) e^{-(\alpha-\rho)t} dt \\
& -\underline{c} \int_0^\infty p_{q^w_{\varphi}}(0,x,t,S_\delta) e^{-(\alpha-\rho)t} dt.
\end{aligned}$$

Since $\underline{c} \leq 0$, and $p_{q^w_{\varphi}}(0,x,t,S_\delta) \leq 1$, it follows that

$$\begin{aligned}
& \int_0^\infty \int_S p_{q^w_{\varphi^*}}(0,x,t,dy) \frac{c_0(y,\varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt \\
\leq & \int_0^\infty \int_S p_{q^w_{\varphi}}(0,x,t,dy) \int_A \frac{c_0(y,a)}{w(y)} \varphi(da|y,t) e^{-(\alpha-\rho)t} dt, \ \forall\ x \in S.
\end{aligned}$$

Condition (16) can be checked using Theorem 2 in the appendix. The similar reasoning also holds for the constrained problem. To avoid repetition, we omit the details.

Appendix A. Some facts about Markov pure jump processes

A (Borel-measurable) signed kernel $q(dy|x,s)$ on $\mathcal{B}(S)$ from $S \times [0,\infty)$ is called a (conservative stable) Q-function on the Borel space S if the following conditions are satisfied.

(a) For each $s \geq 0$, $x \in S$ and $\Gamma \in \mathcal{B}(S)$ with $x \notin \Gamma$, $\infty > q(\Gamma|x,s) \geq 0$.

(b) For each $(x,s) \in S \times [0,\infty)$, $q(S|x,s) = 0$.

(c) For each $x \in S$, $\sup_{s\in[0,\infty)} \{q(S \setminus \{x\}|x,s)\} < \infty$.

For each Q-function q on S, we put $\widetilde{q}(\Gamma|x,s) := q(\Gamma \setminus \{x\}|x,s)$, and $q_x(s) := \widetilde{q}(S|x,s)$.

Given a Q-function q on S from $S \times [0,\infty)$, for each $\Gamma \in \mathcal{B}(S)$, $x \in S$, $s,t \in [0,\infty)$ and $s \leq t$, one can define

$$\begin{aligned}
& p^{(0)}_q(s,x,t,\Gamma) := \delta_x(\Gamma) e^{-\int_s^t q_x(v)dv}, \\
& p^{(n+1)}_q(s,x,t,\Gamma) := \int_s^t e^{-\int_s^u q_x(v)dv} \left(\int_S p^{(n)}_q(u,z,t,\Gamma) \widetilde{q}(dz|x,u) \right) du, \\
& \forall\ n = 0,1,\ldots.
\end{aligned}$$

It is clear that one can legitimately define the sub-stochastic kernel $p_q(s,x,t,dy)$ on S by

$$p_q(s,x,t,\Gamma) := \sum_{n=0}^{\infty} p_q^{(n)}(s,x,t,\Gamma) \tag{17}$$

for each $x \in S$, $s,t \in [0,\infty)$, $s \le t$, and $\Gamma \in \mathcal{B}(S)$. This is the Feller's construction for a transition function, i.e., p_q satisfies

$$p_q(s,x,s,dy) = \delta_x(dy)$$

and the Kolmogorov-Chapman equation

$$\int_S p_q(s,x,t,dy) p_q(t,y,u,\Gamma) = p_q(s,x,u,\Gamma), \ \forall\, \Gamma \in \mathcal{B}(S)$$

is valid for each $0 \le s \le t \le u < \infty$.

Condition 6. *There exist a monotone nondecreasing sequence* $\{\widetilde{V}_n\}_{n=1}^{\infty} \subseteq \mathcal{B}([0,\infty) \times S)$ *and a* $[0,\infty)$*-valued measurable function* $\widetilde{w}$ *on* $[0,\infty) \times S$ *such that the following hold.*

(a) As $n \uparrow \infty$, $\widetilde{V}_n \uparrow [0,\infty) \times S$.

(b) For each $n = 1,2,\ldots,$ $\sup_{x \in \hat{V}_n,\ t \in [0,\infty)} q_x(t) < \infty$, *where* $\hat{V}_n$ *denotes the projection of* $\widetilde{V}_n$ *on* S.

(c) As $n \uparrow \infty$, $\inf_{(t,x) \in ([0,\infty) \times S) \setminus \widetilde{V}_n} \widetilde{w}(t,x) \uparrow \infty$.

(d) For some constant $\rho' \in (0,\infty)$, *for each* $x \in S$ *and* $v \in [0,\infty)$,

$$\int_0^{\infty} \int_S \widetilde{w}(t+v,y) e^{-\rho' t - \int_0^t q_x(s+v)ds} \widetilde{q}(dy|x,t+v)dt \le \widetilde{w}(v,x).$$

The next statement follows from Theorem 3.2 of [41].

Theorem 2. *If Condition 6 is satisfied, then* $p_q(s,x,t,S) = 1$ *for each* $x \in S$, $s,t \in [0,\infty)$ *such that* $s \le t$.

Acknowledgements

This work is partially supported by a grant from the Royal Society (IE160503). We would like to thank the associate editor and referee for very careful reading and useful remarks, which have improved the paper significantly. In particular, the proof technique in Remark 1 was summarized by the referee.

References

[1] Altman, E. (1999). *Constrained Markov Decision Processes*. Chapman and Hall/CRC, Boca Raton.

[2] Anderson, W. (1991). *Continuous-time Markov Chains*. Springer, New York.

[3] Bäuerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer, Berlin.

[4] Blok, H. and Spieksma, F. (2015). Countable state Markov decision processes with unbounded jump rates and discounted optimality equation and approximations. *Adv. Appl. Probab.* **47**, 1088-1107.

[5] Chen, A. Pollett, P. Zhang, H. and Cairns, B. (2005). Uniqueness criteria for continuous-time Markov chains with general transition structures. *Adv. Appl. Probab.* **37**, 1056–1074.

[6] Chen, M. (2015). Practical criterion for uniqueness of q-processes. *Chinese J. Appl. Probab. Stat.*, **31**, 213–224.

[7] Costa, O. and Dufour, F. (2015). A linear programming formulation for constrained discounted continuous control for piecewise deterministic Markov processes. *J. Math. Anal. Appl.* **424**, 892-914.

[8] Dufour, F., Horiguchi, M. and Piunovskiy, A. (2012). The expected total cost criterion for Markov decision processes under constraints: a convex analytic approach. *Adv. Appl. Probab.* **44**, 774–793.

[9] Dufour, F. and Piunovskiy, A. (2013). The expected total cost criterion for Markov decision processes under constraints. *Adv. Appl. Probab.* **45**, 837-859.

[10] Dufour, F. and Prieto-Rumean, T. (2016) Conditions for the solvability of the linear programming formulation for constrained discounted Markov decision processes. *Appl. Math. Optim.* **74**, 27-51.

[11] Dynkin, E. and Yushkevich, A. (1979). *Controlled Markov Processes*. Springer, New York.

[12] Feinberg, E. and Sonin, I. (1996). Notes on equivalent stationary policies in Markov decision processes with total rewards. *Math. Meth. Oper. Res.* **44**, 205-221.

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

[14] Feinberg, E. (2012). Reduction of discounted continuous-time MDPs with unbounded jump and reward rates to discrete-time total-reward MDPs. In *Optimization, Control, and Applications of Stochastic Systems*, Hernandez-Hernandez, D. and Minjarez-Sosa, A. (eds): 77-97, Birkhäuser, Bassel.

[15] Feinberg, E. and Rothblum, U. (2012). Splitting randomized stationary policies in total-reward Markov decision processes. *Math. Oper. Res.* **37**, 129-153.

[16] Feinberg, E., Mandava, M. and Shiryaev, A. (2013). Sufficiency of Markov policies for continuous-time Markov decision processes and solutions of Kolmogorov's forward equation for jump Markov processes. In *Proc. 52nd IEEE CDC*, 5728-5732. Dec, 2013, Florence, Italy.

[17] Feinberg, E., Mandava, M. and Shiryaev, A. (2014). On solutions of Kolmogorov's equations for nonhomogeneous jump Markov processes. *J. Math. Anal. Appl.*, **411**, 261–270.

[18] Feinberg, E., Mandava, M. and Shiryaev, A. (2016). Kolmogorov's equations for jump Markov processes with unbounded jump rates. Available at arXiv:1603.02367.

[19] Gihman, I. and Skorohod, A. (1975). *The Theory of Stochastic Processes II*. Springer, Berlin.

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

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

[22] Guo, X. and Zhang, Y. (2017). Constrained total undiscounted continuous-time Markov decision processes. *Bernoulli* **23**, 1694-1736.

[23] Hernández-Lerma, O. and Lasserre, J. (1996). *Discrete-Time Markov Control Processes*. Springer-Verlag, New York.

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

[25] Jaśkiewicz, A. and Nowak, A. (2011). Stochastic games with unbounded payoffs: applications to robust control in economics. *Dyn. Games Appl.* **1**, 253-279.

[26] Jaśkiewicz, A. and Nowak, A. (2011). Discounted dynamic programming with unbouned returns: application to economic models. *J. Math. Anal. Appl.* **378**, 450-462.

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

[28] Kuznetsov, S. (1984). Inhomogeneous Markov processes. *J. Soviet Math.* **25**, 1380–1498.

[29] Miller, B. (1968). Finite state continuous time Markov decision processes with an infinite planning horizon. *J. Math. Anal. Appl.* **22**, 552–569.

[30] Piunovskiy, A. (1997). *Optimal Control of Random Sequences in Problems with Constraints*, Kluwer, Dordrecht.

[31] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time markov decision processes with unbounded rates: the dynamic programming approach. Available at arXiv:1103.0134.

[32] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim.* **49**, 2032-2061.

[33] Piunovskiy, A. (2015). Randomized and relaxed strategies in continuous-time Markov decision processes. *SIAM J. Control Optim.* **53**, 3503-3533.

[34] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). *Selected Topics in Continuous-Time Controlled Markov Chains and Markov Games*. Imperial College Press, London.

[35] Rykov, V. (1966). Markov decision processes with finite state and decision spaces. *Theory Probab. Appl.* **11**, 302-311.

[36] Spieksma, F. (2015). Countable state Markov processes: non-explosiveness and moment function. *Probab. Eng. Inform. Sc.*, **29**, 623–637.

[37] Spieksma, F. (2016). Kolmogorov forward equation and explosiveness in countable state Markov processes. *Ann. Oper. Res.* **241**, 3–22.

[38] Van, C. and Morhaim, L. (2002). Optimal growth models with bounded or unbounded returns: a unifying approach. *J. Econom. Theory* **105**, 158-187.

[39] Veinott, A. (1969). Discrete dynamic programming with sensitive discount optimality criteria. *Ann. Math. Stat.* **40**, 1635-1660.

[40] van der Wal, J. (1980). *Stochastic Dynamic Programming: Successive Approximations and Nearly Optimal Strategies for Markov Decision Processes and Markov Games*. Mathematisch Centrum, Amsterdam.

[41] Zhang, Y. (2017). On the nonexplosion and explosion for nonhomogeneous Markov pure jump processes. *J. Theor. Probab.*, accepted.