Note on discounted continuous-time Markov decision processes with a lower bounding function

Xin Guo*, Alexey Piunovskiy† and Yi Zhang ‡

Abstract: In this paper, we consider the discounted continuous-time Markov decision process (CTMDP) with a lower bounding function. In this model, the negative part of each cost rate is bounded by the 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 a useful transformation for nonhomogeneous Markov pure jump processes that has not yet been widely applied to the study of CTMDPs. This technique was not employed in previous literature, but it clarifies the roles of the imposed conditions in a rather transparent way. As a consequence, we withdraw and weaken several conditions commonly imposed in the literature.

1 Introduction

In this paper, we consider the discounted continuous-time Markov decision process (CTMDP) with a lower bounding function. In this model, the negative part of each cost rate is bounded by the 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 a useful transformation for nonhomogeneous Markov pure jump processes that has not yet been widely applied to the study of CTMDPs.

Discounted CTMDPs have been studied intensively since the 1960s, with one of the first works being [31]. Initially the 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., [13, 18, 28]. 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 [19, 30] and the articles [4, 28]. If, as in the present paper, we only bound the growth of the negative part of each cost rate using the function w, which is thus called a lower bounding

---

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

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

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

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, thanks to the powerful Feinberg's reduction technique [12, 13], 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. (By the way, Feinberg's reduction technique is different from and much more powerful than the uniformization technique, and its extension to the total undiscounted CTMDP problems is more delicate, see [20, 29].) This approach has been applied to studying the discounted CTMDP problem with arbitrarily unbounded transition rate and nonnegative cost rates, see [13]. 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, and the cost functions can take both positive and negative values. Without additional conditions, the studies for such DTMDPs, especially for constrained problems, are challenging and difficult, as demonstrated in [11], and are still underdeveloped, see e.g., [8].

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, 23], where the motivation for considering this type of cost functions was explained with applications to economics in [23]. Note that they can be reduced to equivalent discounted problems with nonnegative cost functions using the method in [34], see also [9]. 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 nonhomogeneous 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 Feinberg's 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 undiscounted 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 [18, 28] 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. The paper is finished with a conclusion in Section 4. 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 [12, 13, 24, 28], 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 $\widetilde{q}(\Gamma|x, a) := q(\Gamma_S \setminus \{x\}|x, a) \geq 0$ 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 $(x_0, \theta_1, \ldots, \theta_n, x_n, \infty, x_\infty, \infty, x_\infty, \ldots)$, where $x_0, x_1, \ldots, x_n$ belong to S, $\theta_1, \ldots, \theta_n$ 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 $\omega := (x_0, \theta_1, x_1, \theta_2, \ldots) \in \Omega$, 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 [24] 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 $x \in S_\infty$, and below we denote $S_\infty := S \bigcup \{x_\infty\}$.

Definition 2.1 *(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, for each $\omega = (x_0, \theta_1, x_1, \theta_2, \ldots) \in \Omega$ and $t \in (0, \infty)$,*

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

where $a_\infty \notin A$ is some isolated point. Here, for each $n = 0, 1, 2, \ldots$, $\pi_n(da|x_0, \theta_1, \ldots, x_n, s)$ is a stochastic kernel on A concentrated on $A(x_n)$ given $x_0 \in S$, $\theta_1 \in (0, \infty), \ldots$, $x_n \in S$, $s \in (0, \infty)$. We often 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, each of the stochastic kernels π_n reads $\pi_n(da|x_0, \theta_1, \ldots, x_n, s) = \pi(da|x_n)$. 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 by 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

$$\nu^\pi(dt, dy) := \int_A \widetilde{q}(dy|\xi_{t-}(\omega), a)\pi(da|\omega, t)dt$$

$$= \sum_{n \geq 0} \int_A \widetilde{q}(dy|x_n, a)\pi_n(da|x_0, \theta_1, \ldots, \theta_n, x_n, t - t_n)I\{t_n < t \leq t_{n+1}\}dt$$

3

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, i.e., 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 [22, 24]. 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 [22], the conditional distribution of (θ_{n+1}, x_{n+1}) with the condition on $x_0, \theta_1, \ldots, \theta_n, x_n$ is given on $\{\omega : x_n(\omega) \in S\}$ by

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

and given on $\{\omega : x_n(\omega) = x_\infty\}$ 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:

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

and

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

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

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

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

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

4

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

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

for each feasible policy π.

Note that the optimality of a feasible policy for the constrained problem (3) is for the 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 2.1 *There exists a $[1, \infty)$-valued measurable function w on S such that*

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

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

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

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

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

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

Lemma 2.1 *Suppose Condition 2.1 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 $E_x^\pi\left[\int_0^\infty e^{-\alpha t}\int_A c_i(\xi_t, a)\pi(da|\omega, t)dt\right]$, $i = 0, 1, \dots, N$, are well defined.

Proof. This follows from Lemma 2 of [27] and (4). □

Assumption 2.1 *Throughout this paper, unless stated otherwise, Condition 2.1 is assumed to hold automatically, without specific reference.*

3 Main statement and its proof

3.1 Conditions, statements and comments

Condition 3.1 *There exist a $(0, \infty)$-valued measurable function w' on S and a monotone nondecreasing sequence of measurable subsets $\{V_m\}_{m=1}^\infty \subseteq \mathcal{B}(S)$ such that the following hold.*

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

5

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

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

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

(d) $\inf_{x \in S \setminus V_m} \frac{w'(x)}{w(x)} \to \infty$ *as* $m \to \infty$, *where the function* w *comes from Condition 2.1*.

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 3.2   *(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 2.1.*

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

The next condition is for constrained problem only.

Condition 3.3 *There exists a feasible policy for problem (3) with a finite value.*

The main statement of this paper is the following one.

Theorem 3.1 *Suppose Conditions 2.1, 3.1 and 3.2 are satisfied. Then the following assertions hold.*

(a) There exists a deterministic stationary optimal policy for the unconstrained problem (2).

(b) If Condition 3.3 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, 23], 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, 23] was direct. But it is possible to reduce this to equivalent problems with nonnegative cost functions, using the technique in p.101 of [34], see also [9] and p.79 of [1]. The proof of Theorem 3.1 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 2.1, Theorem 3.1(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. Our result here formally shows 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 3.1(b) was obtained in e.g., [28] but under stronger conditions. We include them here for ease of reference.

Instead of Condition 3.1, the following condition was imposed in [28].

Condition 3.4 *There exists a $(0, \infty)$-valued measurable function $\widetilde{w}'$ on S such that the following hold.*

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

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

(c) For some constant $\widetilde{L} \in (0, \infty)$, $(\overline{q}_x + 1) w(x) \leq \widetilde{L} \widetilde{w}'(x)$ for each $x \in S$, where the function w comes from Condition 2.1.

It is easy to see that, if the above condition is satisfied, then so is Condition 3.1 with $w' = \widetilde{w}' + 1$, $\rho' = \widetilde{\rho}'$, $V_m = \left\{ x \in S : \frac{\widetilde{w}'(x) + 1}{w(x)} \leq m \right\}$ for each $m = 1, 2, \ldots$.

Furthermore, under Conditions 2.1, 3.1 and 3.3, in addition to Condition 3.2, it was also assumed in [28] that the function $\frac{\widetilde{w}'}{w}$ is a moment function on $\mathbb{K}$, see Definition E.7 of [21], in order to apply the Prokhorov theorem in their proof, see Proposition E.8 and Theorem E.6 of [21]. This is not needed here. The investigations in [28] 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 3.1. In the way, we comment and clarify the roles of the imposed conditions, and present the auxiliary statements.

3.2 Proof of the main statement

In this subsection, we present the proof of Theorem 3.1, by combining several lemmas. To make the argument as transparent as possible, we proceed our proof in such a way that a lemma is presented only in the place, where it is needed in our proof, instead of collecting them altogether upfront.

Proof of Theorem 3.1. The following statement is a consequence of Theorem 4.2 of [15], and is the starting point of our reasoning.

Lemma 3.1 *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, one can be restricted to the class of Markov policies for problems (2) and (3), i.e., if one obtains 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 [16], see p.262 therein. For the ease of reference, we recall some relevant definitions and facts about Q-functions in the appendix.

According to Theorem 2.2 of [16], 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 [17]. Here and below, we omit the subscript in P_γ^φ, whenever the initial distribution γ is irrelevant. Furthermore, by Theorem 2.2 of [16], p_{q_φ} defined by (16) 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 [25].) Consequently, for each Markov policy φ,

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

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_φ^w 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_φ^w on S_δ is defined by

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

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

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

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. For (uncontrolled) homogeneous continuous-time Markov chains, this transformation was considered in e.g., [2, 32, 33]. But it has not been widely applied to the study of CTMDPs.

Lemma 3.2 *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;*

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

Proof. See Lemma A.3 of [35]. $\qquad \square$

By Lemma 3.2, we see that for each $i = 0, 1, \ldots, N$,

$$w(x) \int_0^\infty \int_S p_{q_\varphi^w}(0, x, t, dy) \int_A \frac{c_i(y, a)}{w(y)} \varphi(da|y, t) e^{-(\alpha-\rho)t} dt$$

$$= \int_0^\infty \int_S \int_A c_i(y, a) \varphi(da|y, t) e^{-\alpha t} p_{q_\varphi}(0, x, t, dy) dt, \ \forall \, x \in S.$$

Hence, problem (2) is equivalent to

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

8

and problem (3) is equivalent to

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

$$\text{such that} \quad \int_0^\infty \int_S p_{q_\varphi^w}(0, x, t, dy) \int_A \frac{c_j(y, a)}{w(y)} \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \leq \frac{d_j}{w(x)},$$

$$j = 1, 2, \ldots, N. \tag{7}$$

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

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

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

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

The requirement of $\alpha > \rho$ in Condition 2.1(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_i^w defined by

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

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

$$c_i^w(\delta, a_\infty) := 0.$$

According to the Feinberg's reduction technique for discounted CTMDPs, see [13], 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

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

9

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 $i = 0, 1, \ldots, N$, 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., [21, 26] 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, and the cost function C_i can take both positive and negative values. (This is the case e.g., if the original CTMDP is an uncontrolled pure birth process with $S = \{1, 2, \ldots\}$, and birth rate at the state $x \in S$ being $2x$, $\alpha = 2$, $\rho = 1$ and $w(x) = 1$ for each $x \in S$.) Compared to the absorbing model treated in [1, 14], the theory for such a DTMDP model is technical and demanding, and, without additional assumptions, there is far less result concerning the existence of stationary strategies, which one can directly refer to, especially for the constrained problems, see [8, 11].

On the other hand, the functions c_i^w, $i = 0, 1, \ldots, N$, are bounded from below under Condition 2.1(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 $i = 0, 1, \ldots, N$. Then the functions $\widetilde{c}_i^w$ are all nonnegative. In order for problems (6) and (7) to be equivalent to

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

and

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

$$\text{such that} \quad \int_0^\infty \int_{S_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{A_\delta} \widetilde{c}_j^w(y) \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \le \frac{d_j}{w(x)} - \frac{\underline{c}}{\alpha - \rho},$$

$$j = 1, 2, \ldots, N, \tag{10}$$

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

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

Condition 3.1 is precisely imposed for this purpose, as seen in the next statement. (An alternative justification of the role of Condition 3.1 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 justification here is more transparent in our opinion.)

Lemma 3.3 *Let some Markov policy φ be fixed. Suppose Condition 2.1(a) and Condition 3.1 are satisfied. Then (11) holds.*

Proof. According to Theorem A.1, for the statement it suffices to verify that Condition A.1 is satisfied.

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

$$\int_S \frac{w'(y)}{w(y)} q^w(dy|x,s) = \int_S \frac{w'(y)}{w(y)} \frac{w(y)}{w(x)} \widetilde{q}(dy|x,s) - (\rho + q_x(s)) \frac{w'(x)}{w(x)}$$

$$= \int_S \frac{w'(y)}{w(x)} \widetilde{q}(dy|x,s) - (\rho + q_x(s)) \frac{w'(x)}{w(x)} \leq (\rho' - \rho) \frac{w'(x)}{w(x)}, \ \forall \ x \in S, \ s \geq 0. \tag{12}$$

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 A.1, 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 A.1 is satisfied because, by (12),

$$\int_0^\infty \int_{S_\delta} \widetilde{w}(t+v,y) e^{-\rho' t - \int_{(0,t]} q_x^w(s+v)ds} \widetilde{q^w}(dy|x,t+v)dt$$

$$\leq \int_0^\infty e^{-\rho' t - \int_0^t q_x^w(s+v)ds} \left(q_x(s) + \rho' \right) \widetilde{w}(v,x) = \widetilde{w}(v,x), \ \forall \ x \in S,$$

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

Thus, by Theorem A.1, we see that relation (11) is satisfied, and the statement follows. $\square$

By the way, under Condition 2.1(a), in certain models, Condition 3.1 is also necessary for (11) to hold under certain policies; see [35]. In the homogeneous denumerable case, this was first observed in [32]. For more concrete examples such as single birth processes, this necessity part was known earlier, see [5].

As a result of the above lemma and the discussions above it, we see that under Condition 2.1 and Condition 3.1, 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 Feinberg's reduction technique [13], 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

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

$$\text{such that} \qquad \mathbb{E}_x^\sigma \left[\sum_{n=0}^\infty \widetilde{C}_j(X_n, A_n) \right] \leq \frac{d_j}{w(x)} - \frac{c}{\alpha - \rho},$$

$$j = 1, 2, \ldots, N, \tag{14}$$

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

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

11

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 2.1, Condition 3.1, and Condition 3.2 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 semi-continuous model, see [3, 10], 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 3.3, one can refer to Theorem 4.1 of [7], see also Theorem A.2 of [6], 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 Feinberg's 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). □

We finish this section with the following remark. In general, problems (6) and (7) are not equivalent to (9) and (10), respectively. According to [13], (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 2.1 and Condition 3.2, 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:

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

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.,

$$p_{q_{\varphi^*}^w}(0, x, t, S_\delta) = 1, \ \forall \ x \in S, \ t \in [0, \infty). \tag{15}$$

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),

$$\inf_{\varphi \in \Pi^M} \left\{ \int_0^\infty \int_{S_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{A_\delta} \widetilde{c}_0^w(y, a) \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \right\}$$

$$= \int_0^\infty \int_{S_\delta} p_{q_{\varphi^*}^w}(0, x, t, dy) \widetilde{c}_0^w(y, \varphi^*(y)) e^{-(\alpha-\rho)t} dt$$

$$= \int_0^\infty \int_S p_{q_{\varphi^*}^w}(0, x, t, dy) \frac{c_0(y, \varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt - \frac{\underline{c}}{\alpha-\rho}, \ \forall \ x \in S.$$

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

$$\int_0^\infty \int_S p_{q_{\varphi^*}^w}(0, x, t, dy) \frac{c_0(y, \varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt - \frac{\underline{c}}{\alpha-\rho}$$

$$\leq \int_0^\infty \int_{S_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{A_\delta} \widetilde{c}_0^w(y, a) \varphi(da|y, t) e^{-(\alpha-\rho)t} dt$$

$$= \int_0^\infty \int_S p_{q_\varphi^w}(0, x, t, dy) \int_A \frac{c_0(y, a)}{w(y)} \varphi(da|y, t) e^{-(\alpha-\rho)t} dt - \underline{c} \int_0^\infty p_{q_\varphi^w}(0, x, t, S_\delta) e^{-(\alpha-\rho)t} dt.$$

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

$$\int_0^\infty \int_S p_{q_{\varphi^*}^w}(0, x, t, dy) \frac{c_0(y, \varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt$$

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

Condition (15) can be checked using Theorem A.1 in the appendix. The similar reasoning also holds for the constrained problem. To avoid repetition, we omit the details.

4 Conclusion

To sum up, we showed the existence of a deterministic stationary (respectively, stationary) optimal policy for the unconstrained (respectively, constrained) discounted CTMDP problems under rather weak conditions. The main feature in the model is that only the negative part of each cost rate is bounded by a drift function. Another contribution is that our arguments were based on a transformation for Markov pure jump processes, and this technique had not been widely applied to the study of CTMDPs. On the other hand, exactly this technique allowed us to clarify the roles of all the imposed conditions in a transparent way. In this way, even in the specific case, where both the negative and positive parts of the cost rates are bounded by the drift function, we improved the existing results in the literature by withdrawing several and various conditions assumed therein.

A Appendix

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

$$p_q^{(0)}(s, x, t, \Gamma) := \delta_x(\Gamma) e^{-\int_s^t q_x(v) dv},$$

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

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

$$p_q(s, x, t, \Gamma) := \sum_{n=0}^\infty p_q^{(n)}(s, x, t, \Gamma) \tag{16}$$

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

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

13

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 A.1 *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 [35].

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

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

[6] 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.

[7] 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.

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

[9] 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.

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

[11] 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.

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

[13] 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.

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

[15] 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.

[16] 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.

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

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

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

[20] Guo, X. and Zhang, Y. (2016). Constrained total undiscounted continuous-time Markov decision processes. *Bernoulli*, accepted.

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

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

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

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

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

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

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

[28] 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.

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

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

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

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

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

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

[35] Zhang, Y. (2015). On the nonexplosion and explosion for nonhomogeneous Markov pure jump processes. Preprint. Available at arXiv:http://arxiv.org/abs/1511.05011.