Bernoulli **23**(3), 2017, 1694–1736
DOI: 10.3150/15-BEJ793

Constrained total undiscounted continuous-time Markov decision processes

XIANPING GUO[1] and YI ZHANG[2]

[1]*School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, P.R. China.*
E-mail: mcsgxp@mail.sysu.edu.cn
[2]*Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.*
E-mail: yi.zhang@liv.ac.uk

The present paper considers the constrained optimal control problem with total undiscounted criteria for a continuous-time Markov decision process (CTMDP) in Borel state and action spaces. The cost rates are nonnegative. Under the standard compactness and continuity conditions, we show the existence of an optimal stationary policy out of the class of general nonstationary ones. In the process, we justify the reduction of the CTMDP model to a discrete-time Markov decision process (DTMDP) model based on the studies of the undiscounted occupancy and occupation measures. We allow that the controlled process is not necessarily absorbing, and the transition rates are not necessarily separated from zero, and can be arbitrarily unbounded; these features count for the main technical difficulties in studying undiscounted CTMDP models.

Keywords: constrained optimality; continuous-time Markov decision processes; total undiscounted criteria

1. Introduction

The present paper considers the constrained optimal control problem with total undiscounted criteria for a continuous-time Markov decision process (CTMDP) in Borel state and action spaces. The cost rates are nonnegative.

The majority of the previous literature on CTMDPs with the total cost criteria focuses on the discounted model with a positive constant discount factor; see, for example, [13,14,19–21, 27,29–31,33,36]. In [21,29,30], the convex analytic approach for constrained problems is developed, whereas the dynamic programming approach for unconstrained problems is studied in [19,20,31,33]. The investigations in [19–21,29,30,33] are based on the direct investigation of the continuous-time models by using the Kolmogorov forward equations; for this, the authors had to impose extra conditions bounding the growth of the transition rates in the form of the existence of Lyapunov functions.

Another method of investigation is based on the study of the relation of the CTMDP problem and a DTMDP (discrete-time Markov decision process) problem. Once the CTMDP problem is reduced to an equivalent DTMDP problem, one can directly make use of the toolbox of the better developed theory of DTMDPs; see, for example, [2,4,5,10,12,17,32] for the CTMDPs. This idea at least dates back to the 1970s; see [28], where the author applied the uniformization technique to reducing the CTMDP problem to a DTMDP problem; see also [36]. However, the authors of [28,36], not only required the transition rates to be uniformly bounded, also had to be restricted to the class of deterministic stationary policies, that is, those that do not change actions

between two consecutive state transitions. These are also the standard setup for the textbook treatment of CTMDPs; see Chapter 11 of [34]. The situation becomes more complicated if one is allowed, as in the present paper, to consider nonstationary policies, that is, those allowing the change in actions between two state transitions. In this direction, Yushkevich [37] firstly reduced a CTMDP model with nonstationary policies to a DTMDP model, which, for the ease of reference, we call the Yushkevich-induced model. However, the action space of the Yushkevich-induced DTMDP model is more complicated; it is the space of measurable mappings, so that in general a stationary policy in this DTMDP model corresponds to a nonstationry policy for the original CTMDP model. A further reduction of the Yushkevich-induced DTMDP model to one with the same action space as the original CTMDP model is possible after the investigations of the dynamic programming (or say optimality) equation for unconstrained problems. Only unconstrained problems were considered in [37], which also assumed the transition rates in the CTMDP model to be uniformly bounded. The similar approach is developed in [3,7,8,18] for other related unconstrained problems.

In general, the reduction method based on the comparison of the dynamic programming equation is more suitable for unconstrained problem; see also [31]. Especially convenient for dealing with constrained discounted CTMDP problems, [13,14] proposed a novel method of reducing directly the CTMDP model to an equivalent DTMDP model in the same action space based on the studies of the discounted occupancy measures; we often call such an induced DTMDP model "simple" to distinguish it from the Yushkevich-induced DTMDP model. (In fact, there is a small inconsistency in the use of terminologies in [13,14]; the occupation measure in [13] actually means the occupancy measure in [14] as well as the present paper.) The original article [13] assumed the transition rates of the CTMDP to be bounded; this condition is completely withdrawn in the more recent extension [14]. Feinberg's reduction is valid without any conditions so long the discount factor is positive. By the way, in [13,14], the author considered general cost rates, whereas in the present paper we only consider nonnegative cost rates.

The present paper considers the total undiscounted CTMDP problem with constraints. To the best of our knowledge, the theory for this class of optimal control problems is currently underdeveloped, despite that they would naturally find applications to for example, epidemiology, where one aims at minimizing the total endemic time, which does not have an obvious monetary interpretation for discounting. There seems to be limited literature on this topic. For unconstrained total undiscounted problem, Forwick, Schäl and Schmitz [18] developed the dynamic programming approach, and established the optimality equation, essentially following the Yushevich's reduction method. For the constrained problem, the authors of [22] developed the convex analytic approach by studying directly the continuous-time model, but only after imposing the extra conditions on the growth of the transition and cost rates and some strongly absorbing structure; such conditions make the analysis of the undiscounted model essentially similar to the one of a discounted one.

The objective of this paper is to study the constrained total undiscounted CTMDP problem without the absorbing condition or any condition on the growth of the transition rate, whereas the cost rates are nonnegative. Even for DTMDPs, such problems were acknowledged to be challenging in the survey [6] and were tackled only recently in [10]; see also [11,12]. Our original plan is to apply the Feinberg's reduction method to the undiscounted case, and once that is done, we can refer to the optimality results for DTMDPs obtained in [10]; we remark that the Feinberg's reduction method is always applicable to discounted CTMDP models without additional

conditions. However, we notice that the situation when the discount factor for the CTMDP model is zero becomes significantly different and much more delicate; indeed, Example 3.1 below illustrates that without additional conditions (in fact, when the transition rate is not separated from zero), it can happen that the performance vector of the CTMDP problem under a nonstationary policy might not be replicated by any performance vectors of the simple-induced DTMDP problem. It is thus natural to ask under what conditions does the reduction (to the simple-induced DTMDP model) method apply to the undiscounted CTMDP model. It is also realized that the studies of the occupancy measures alone are not useful in general for the total undiscounted CTMDP models. (In Section 3 below, we give a more detailed discussion on these.) Different from the discounted case, we now also need study the occupation measures, which are on the one hand, more delicate because they are infinitely valued, and on the other hand, are more suitable and convenient for constrained problems. (In particular, they were not considered in [37] or [18] dealing with unconstrained problems.)

Having said the above, the main contributions of the present paper are as follows. (More detailed comments on the novelty and contributions of the paper are postponed to the end of Section 3.)

(a) We provide the natural condition for the validity of reducing the total undiscounted CTMDP model with constraints to a simple-induced DTMDP model. Our conditions are of the standard continuity and compactness type, and allow the transition rates not necessarily separated from zero on the one hand, and arbitrarily unbounded on the other hand. No absorbing structure is assumed. The approach in [22] are not applicable in this general setup. Also note that the arguments in [13,14] are essentially based on the presence of the positive discount factor; see Section 3 for greater details.

(b) We show the existence of an optimal stationary policy out of the class of general (nonstationary) ones. It is arguable that the solvability, as we confine ourselves to in this paper, is an issue of core importance to be addressed first for any optimal control problem. In the present general setup, it is not clear how to obtain this result following the widely used method in the literature based on the Dynkin's formula.

(c) The paper is not a simple extension of the uniformization technique for CTMDPs, as explained in the above. Rather, our investigations are based on the delicate studies of undiscounted occupancy measures and occupation measures of the CTMDP model, for which we incidentally obtain some properties of independent interest.

The rest of this paper is organized as follows. We describe the controlled process and state the concerned optimal control problems in Section 2. In Section 3, we provide some relevant facts about discounted CTMDPs, and an example further demonstrating the contribution and novelty of the present paper. In Sections 4 and 5, we obtain some properties of the occupancy and occupation measures, respectively. In Section 6, we establish the optimality results. We end this paper with a conclusion in Section 7.

2. Optimal control problem statement

The objective of this section is to describe briefly the controlled process similarly to [26,27,29], and the associated optimal control problem of interest in this paper.

Notations and conventions. In what follows, I stands for the indicator function, $\delta_x(\cdot)$ is the Dirac measure concentrated at x, and $\mathcal{B}(X)$ is the Borel σ-algebra of the topological space X. A measure is σ-additive and $[0,\infty]$-valued. The abbreviation s.t. (resp., a.s.) stands for "subject to" (resp., "almost surely"). Below, unless stated otherwise, the term of measurability is always understood in the Borel sense. Throughout this article, we adopt the conventions of $\frac{0}{0} := 0$, $0 \cdot \infty := 0$ and $\frac{1}{0} := +\infty$.

2.1. Description of the CTMDP

The primitives of a CTMDP model are the following elements $\{S, A, q, \gamma\}$, where S is a nonempty Borel state space, A is a nonempty Borel action space, γ is a probability measure on $\mathcal{B}(S)$ and represents the initial distribution, and q stands for a signed kernel $q(dy|x,a)$ on $\mathcal{B}(S)$ given $(x,a) \in S \times A$ such that $\tilde{q}(\Gamma_S|x,a) := q(\Gamma_S \setminus \{x\}|x,a) \geq 0$ for all $\Gamma_S \in \mathcal{B}(S)$. Throughout this article we assume that $q(\cdot|x,a)$ is conservative and stable, i.e., $q(S|x,a) = 0$ and $\bar{q}_x = \sup_{a\in A(x)} q_x(a) < \infty$, where $q_x(a) := -q(\{x\}|x,a)$. The signed kernel q is often called the transition rate. Throughout this article, $\bar{q}_x$ is allowed to be arbitrarily unbounded in $x \in S$, unlike in [19,22,29,30]. In line with [10,18] and to fix ideas, we do not consider the case of different admissible action spaces at different states.

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 the limit point of the sequence $\{t_n\}$ is denoted by $t_\infty(\omega) := \lim_{n\to\infty} t_n(\omega)$. 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

$$\xi_t = \sum_{n\geq 0} I\{t_n \leq t < t_{n+1}\} x_n + I\{t_\infty \leq t\} x_\infty; \tag{2.1}$$

recall that x_∞ is the isolated point. Below we denote $S_\infty := S \cup \{x_\infty\}$, and accept $0 \cdot x := 0$ and $1 \cdot x := x$ for each $x \in S_\infty$.

Definition 2.1. *A (history-dependent) policy π for the CTMDP is given by a sequence (π_n) such that, for each $n = 1, 2, \ldots$, $\pi_n(da|x_0, \theta_1, \ldots, x_{n-1}, s)$ is a stochastic kernel on A, and for each $\omega = (x_0, \theta_1, x_1, \theta_2, \ldots) \in \Omega$, $t > 0$,*

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

where $a_\infty \notin A$ is some isolated point. A policy $\pi = (\pi_n)$ is called Markov if, with slight abuse of notations, each of the stochastic kernels π_n reads $\pi_n(da|x_0, \ldots, x_{n-1}, s) = \pi_n(da|x_{n-1}, s)$. A Markov policy is further called deterministic if the stochastic kernels $\pi_n(da|x_{n-1}, s)$ all degenerate. A policy $\pi = (\pi_n)$ is called stationary if, with slight abuse of notations, each of the stochastic kernels π_n reads $\pi_n(da|x_0, \ldots, x_{n-1}, s) = \pi(da|x_{n-1})$. A stationary policy is further called deterministic if $\pi_n(da|x_0, \ldots, x_{n-1}, s) = \delta_{f(x_{n-1})}(da)$ for some measurable mapping f from S to A.

For each $n = 0, 1, \ldots,$ we formally put

$$\pi_{n+1}\big(\{a_\infty\}|x_0, \ldots, x_n, \infty\big) := 1 =: \pi_{n+1}\big(\{a_\infty\}|x_0, \ldots, x_\infty, \infty\big)$$

with $a_\infty \notin A$ being the isolated point.

The class of all policies for the CTMDP model is denoted by Π, and the class of all deterministic Markov policies for the CTMDP model is denoted by Π_{DM}.

Under a policy $\pi := (\pi_n) \in \Pi$, we define the following random measure on $S \times (0, \infty)$

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

with $q_{x_\infty}(a_\infty) = q(dy|x_\infty, a_\infty) := 0$. Then 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 [25,27]. The process $\{\xi_t\}$ defined by (2.1) under the probability measure P_γ^π is called a CTMDP. Below, when $\gamma(\cdot)$ 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.

In what follows, when it is not necessary to emphasize the initial distribution γ, we also say that $\{S, A, q\}$ is our CTMDP model.

2.2. Description of the concerned optimal control problem

Let $N \in \{1, 2, \ldots\}$ be fixed. Consider the nonnegative measurable functions $c_i(x, a) \geq 0$ with $i = 0, 1, \ldots, N$ from $S \times A$ to $[0, \infty)$ as the cost rates. We formally put $c_i(x_\infty, a) := 0$ for each $i = 0, 1, 2, \ldots, N$.

In this paper, we study the following optimal control problem:

$$\begin{aligned} &E_\gamma^\pi\left[\int_0^\infty \int_A c_0(\xi_t,a)\pi(da|\omega,t)\,dt\right] \to \min_{\pi\in\Pi} \\ &\quad \text{s.t.} \quad E_\gamma^\pi\left[\int_0^\infty \int_A c_j(\xi_t,a)\pi(da|\omega,t)\,dt\right] \le d_j, \qquad \forall j=1,2,\dots,N, \end{aligned} \tag{2.2}$$

where for each $j=1,2,\dots,N$, $d_j\in[0,\infty)$ is the fixed constraint constant.

A policy $\pi\in\Pi$ is called feasible for problem (2.2) if

$$E_\gamma^\pi\left[\int_0^\infty \int_A c_j(\xi_t,a)\pi(da|\omega,t)\,dt\right] \le d_j, \qquad j=1,2,\dots,N.$$

Let Π_F be the class of feasible policies. Then the value of problem (2.2) is denoted as

$$V_c(\gamma) := \inf_{\pi\in\Pi_F} E_\gamma^\pi\left[\int_0^\infty \int_A c_0(\xi_t,a)\pi(da|\omega,t)\,dt\right].$$

A feasible policy π for problem (2.2) is called to be with a finite value if

$$E_\gamma^\pi\left[\int_0^\infty \int_A c_0(\xi_t,a)\pi(da|\omega,t)\,dt\right] < \infty.$$

Finally, a policy $\pi^*\in\Pi_F$ is called optimal for the (constrained) CTMDP problem (2.2) if it holds that

$$\inf_{\pi\in\Pi_F} E_\gamma^\pi\left[\int_0^\infty \int_A c_0(\xi_t,a)\pi(da|\omega,t)\,dt\right] = E_\gamma^{\pi^*}\left[\int_0^\infty \int_A c_0(\xi_t,a)\pi^*(da|\omega,t)\,dt\right].$$

3. Facts about the discounted CTMDP problem and discussions

The purpose of this section is to (a) present some relevant results about the α-discounted problem for the CTMDP model $\{S,A,q,\gamma\}$, which are used in the subsequent investigations for our undiscounted CTMDP problem (2.2); and (b) demonstrate the significant difference between the discounted and the undiscounted CTMDP problems, and illustrate that the undiscounted problem is more delicate, which thus clarifies the contribution of the present paper; see Example 3.1 and the discussion following it.

In his well-written articles [13,14], Professor Feinberg considered the following constrained discounted optimal control problem for the CTMDP model $\{S,A,q,\gamma\}$

$$\begin{aligned} &E_\gamma^\pi\left[\int_0^\infty e^{-\alpha t}\int_A c_0(\xi_t,a)\pi(da|\omega,t)\,dt\right] \to \min_{\pi\in\Pi} \\ &\quad \text{s.t.} \quad E_\gamma^\pi\left[\int_0^\infty e^{-\alpha t}\int_A c_j(\xi_t,a)\pi(da|\omega,t)\,dt\right] \le d_j, \qquad \forall j=1,2,\dots,N, \end{aligned} \tag{3.1}$$

where $d_j \in \mathbb{R}$ for each $j = 1, 2, \ldots, N$, and the finite constant $\alpha > 0$ is a fixed discount factor. The investigations in [13,14] are based on the study of the so-called α-discounted occupancy measures, first introduced therein, which we recall as follows.

Definition 3.1. *For each $n = 0, 1, \ldots,$ and (finite) constant $\alpha > 0$, the α-discounted occupancy measure of the policy $\pi \in \Pi$ for the CTMDP model $\{S, A, q, \gamma\}$ is a measure $M_{\gamma,\alpha}^{n,\pi}$ on $\mathcal{B}(S \times A)$ defined by*

$$M_{\gamma,\alpha}^{n,\pi}(\Gamma_S \times \Gamma_A) := E_\gamma^\pi \left[\int_{t_n}^{t_{n+1}} e^{-\alpha t} I\{\xi_t \in \Gamma_S\} \int_{\Gamma_A} \big(\alpha + q_{\xi_t}(a)\big)\pi(da|\omega, t)\, dt \right]$$

for each $\Gamma_S \in \mathcal{B}(S)$ and $\Gamma_A \in \mathcal{B}(A)$.

Professor Feinberg noticed that there is a close relationship between the (α-discounted) occupancy measure for the CTMDP model $\{S, A, q, \gamma\}$ and the marginal distribution of (X_n, A_{n+1}) of the DTMDP model $\{S_\infty, A, p_\alpha, \gamma\}$, where the transition probability p_α is defined for each $\Gamma_S \in \mathcal{B}(S)$ by

$$p_\alpha(\Gamma_S|x, a) = \frac{\tilde{q}(\Gamma_S|x, a)}{\alpha + q_x(a)}, \qquad \forall x \in S, a \in A$$

and

$$p_\alpha(\Gamma_S|x_\infty, a) = 0, \qquad \forall a \in A.$$

Recall that $S_\infty = S \cup \{x_\infty\}$ with $x_\infty \notin S$ being the isolated point. Under each policy σ for the DTMDP model $\{S_\infty, A, p_\alpha, \gamma\}$, let the corresponding strategic measure be denoted by $\mathbf{P}_\gamma^{\alpha,\sigma}$. The expectation taken with respect to $\mathbf{P}_\gamma^{\alpha,\sigma}$ is written as $\mathbf{E}_\gamma^{\alpha,\sigma}$.

The next statement is established in [14].

Proposition 3.1. *The following assertions hold for each $\Gamma_S \in \mathcal{B}(S)$ and $\Gamma_A \in \mathcal{B}(A)$.*

(a) *For each policy $\pi \in \Pi$ for the CTMDP model, there is a Markov policy σ^M for the DTMDP model $\{S_\infty, A, p_\alpha\}$ such that*

$$M_{\gamma,\alpha}^{n,\pi}(\Gamma_S \times \Gamma_A) = \mathbf{P}_\gamma^{\alpha,\sigma^M}(X_n \in \Gamma_S, A_{n+1} \in \Gamma_A), \qquad \forall n = 0, 1, \ldots.$$

(b) *For each Markov policy σ^M for the DTMDP model $\{S_\infty, A, p_\alpha\}$, there exists a Markov policy π^M for the CTMDP model such that*

$$M_{\gamma,\alpha}^{n,\pi^M}(\Gamma_S \times \Gamma_A) = \mathbf{P}_\gamma^{\alpha,\sigma^M}(X_n \in \Gamma_S, A_{n+1} \in \Gamma_A), \qquad \forall n = 0, 1, \ldots.$$

According to Proposition 3.1 and the well-known Derman–Strauch lemma, see [9] or Lemma 2 of [32], for each $\pi \in \Pi$ for the CTMDP model $\{S, A, q, \gamma\}$, there exists some policy σ for the

DTMDP model $\{S_\infty, A, p_\alpha, \gamma\}$ such that

$$\sum_{n=0}^{\infty} M_{\gamma,\alpha}^{n,\pi}(dx \times da) = \sum_{n=0}^{\infty} \mathbf{P}_\gamma^{\alpha,\sigma}(X_n \in dx, A_{n+1} \in da), \tag{3.2}$$

and vice versa. Consequently, the α-discounted CTMDP problem (3.1) can be reduced to the following DTMDP problem for the model $\{S_\infty, A, p_\alpha, \gamma\}$

$$\begin{aligned} &\mathbf{E}_\gamma^{\alpha,\sigma}\left[\sum_{n=0}^{\infty} \frac{c_0(X_n, A_{n+1})}{\alpha + q_{X_n}(A_{n+1})}\right] \to \min_\sigma \\ &\text{s.t.} \quad \mathbf{E}_\gamma^{\alpha,\sigma}\left[\sum_{n=0}^{\infty} \frac{c_j(X_n, A_{n+1})}{\alpha + q_{X_n}(A_{n+1})}\right] \le d_j, \qquad j = 1, 2, \ldots, N. \end{aligned}$$

(Recall that $c_i(x_\infty, a) := 0$ for each $a \in A$.)

Here and below by reduction is meant that both problems have the same value, and if an optimal policy exists for one problem, so does an optimal policy for the other problem.

We emphasize that this reduction for the α-discounted CTMDP problem is possible without any extra conditions being imposed on the CTMDP model, so long $\alpha > 0$.

It is natural to ask whether the reduction is possible for the case of $\alpha = 0$; that is, whether the CTMDP problem (2.2) can be reduced to the following problem

$$\begin{aligned} &\mathbf{E}_\gamma^{\sigma}\left[\sum_{n=0}^{\infty} \frac{c_0(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] \to \min_\sigma \\ &\text{s.t.} \quad \mathbf{E}_\gamma^{\sigma}\left[\sum_{n=0}^{\infty} \frac{c_j(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] \le d_j, \qquad \forall j = 1, 2, \ldots, N \end{aligned} \tag{3.3}$$

for the DTMDP model $\{S_\infty, A, p, \gamma\}$, where the transition probability p being defined for each $\Gamma_S \in \mathcal{B}(S)$ by

$$p(\Gamma_S | x, a) = \frac{\tilde{q}(\Gamma_S | x, a)}{q_x(a)}, \qquad \forall x \in S, a \in A \tag{3.4}$$

and

$$p(\Gamma_S | x_\infty, a) = 0. \tag{3.5}$$

(Recall that $\frac{0}{0} := 0$.) As before, the controlled and controlling processes for the DTMDP model $\{S_\infty, A, p, \gamma\}$ are denoted by $\{X_n\}$ and $\{A_n\}$; $\mathbf{P}_\gamma^\pi$ denotes the strategic measure under the policy σ for this DTMDP model with the corresponding expectation $\mathbf{E}_\gamma^\sigma$.

We remark that since $c_i(x_\infty, a) = 0$ and $p(dy | x_\infty, a) = \delta_{x_\infty}(dy)$ for each $a \in A$, the definition of a policy σ at the current state x_∞ for the DTMDP model $\{S_\infty, A, p, \gamma\}$ is not important for its performance as far as problem (3.3) is concerned, and so we do not specify it in what follows.

The next example shows that the answer to the question mentioned earlier is negative in general.

Example 3.1. Consider the CTMDP model with $S = \{1, 2\}$, $A = [0, \infty)$, $q_1(a) = q(\{2\}|1, a) = e^{-a}$, $q_2(a) = 0$ for each $a \in A$, and $\gamma(\{1\}) = 1$. Let $N = 1$, and $c_0(1, a) = e^{-a}$, $c_0(2, a) = 0$ for each $a \in A$, and $c_1(x, a) = 0$ for each $x \in S$ and $a \in A$. Let $d_1 > 0$, so that any policy is feasible for the CTMDP problem (2.2). Let us fix a policy π defined by

$$\pi\big(\{a\}|\omega, t\big) = \pi_0\big(\{a\}|x, t\big) = I\{a = t\}, \qquad \forall a \in A,$$

so that

$$\int_A q_1(a)\pi(da|1, t) = \int_A c_0(1, a)\pi(da|1, t) = e^{-t}.$$

Then under this policy π, we see

$$E_\gamma^\pi\left[\int_0^\infty \int_A c_0(\xi_t, a)\pi(da|\omega, t)\, dt\right] = E_\gamma^\pi\left[\int_0^{\theta_1} e^{-t}\, dt\right] < \int_0^\infty e^{-t}\, dt = 1,$$

where the third equality is due to the fact $P_\gamma^\pi(\theta_1 = \infty) = e^{-1} < 1$. On the other hand, since $\frac{c_0(1,a)}{q_1(a)} = 1$ for each $a \in A$, we have that under each policy σ for the DTMDP model $\{S_\infty, A, p, \gamma\}$

$$\mathbf{E}_\gamma^\sigma\left[\sum_{n=0}^\infty \frac{c_0(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] = 1.$$

In summary, each policy for the DTMDP $\{S_\infty, A, p, \gamma\}$ model would be optimal for problem (3.3) with the (optimal) value being 1, whereas the value for the CTMDP problem (2.2) is strictly smaller than 1. Hence, the CTMDP problem (2.2) cannot be reduced to the DTMDP problem (3.3).

It is also clear that Proposition 3.1 does not hold in general when $\alpha = 0$.

An objective of the present paper is to provide weak and natural conditions under which the reduction of the CTMDP problem (2.2) to the DTMDP problem (3.3) is possible. To this end, apart from studying the (undiscounted) occupancy measures (see Definition 4.1), we also need investigate the (undiscounted) occupation measures (see Definition 5.1) for the CTMDP model $\{S, A, q, \gamma\}$, for which some properties are to be obtained. The occupation measure is more delicate for studies because it is infinitely valued, whereas the occupancy measure is always finite; see (4.16) below. Finally, under our conditions, we obtain the existence of an optimal stationary policy for the CTMDP problem (2.2). It is arguable that the solvability, as we confine ourselves to in this paper, is an issue of core importance to be addressed for any optimal control problem.

4. Occupancy measure

The objective in this section is to obtain a partial version of Proposition 3.1(a); see Theorem 4.1 below. This statement is needed in the subsequent sections. To this end, we need first impose the compactness-continuity condition; see Condition 4.1 below.

4.1. Compactness-continuity condition and discussions

We introduce another notation before stating the next compactness-continuity condition. Let the stochastic kernel $\tilde{p}$ on $\mathcal{B}(S)$ from $S \times A$ be defined by

$$\tilde{p}(\Gamma|x,a) := \frac{\tilde{q}(\Gamma|x,a)}{q_x(a)} I\{q_x(a) > 0\} + \delta_x(\Gamma) I\{q_x(a) = 0\}, \qquad \forall \Gamma \in \mathcal{B}(S). \tag{4.1}$$

Condition 4.1. (a) *The space A is compact.*

(b) *For each bounded continuous function f on S, $\int_S f(y)\tilde{p}(dy|x,a)$ is continuous in $(x,a) \in S \times A$.*

(c) *$q_x(a)$ is continuous in $(x,a) \in S \times A$.*

(d) *For each $i = 0, 1, \ldots, N$, c_i is lower semicontinuous in $(x,a) \in S \times A$.*

A direct consequence of Condition 4.1 is the next lemma, whose proof is routine and omitted.

Lemma 4.1. *Suppose Condition 4.1(b, c) is satisfied. Then the following assertions hold.*

(a) *For each $[0,\infty]$-valued lower semicontinuous function c on $S \times A$, $\frac{c(x,a)}{q_x(a)}$ is lower semicontinuous in $(x,a) \in S \times A$; and $\int_S f(y)q(dy|x,a)$ is continuous in $(x,a) \in S \times A$ for each bounded continuous function f on S.*

(b) *For each $[0,\infty]$-valued lower semicontinuous function f on S, $\int_S f(y)\tilde{q}(dy|x,a) \in [0,\infty]$ is lower semicontinuous on $S \times A$.*

The assertions in the above lemma are often used without special reference below.

In the original version of this paper, instead of Condition 4.1(b), the following condition is imposed.

Condition 4.2. *For each bounded continuous function $f(x)$ on S, $\int_S f(y)\frac{\tilde{q}(dy|x,a)}{q_x(a)}$ is continuous in $(x,a) \in S \times A$.*

(Condition 4.2 is included here only for the sake of comparisons; it will not be referred to at all in the forthcoming sections.)

Two referees pointed out a flaw of Condition 4.2. Fix $(x,a) \in S \times A$ such that $q_x(a) = 0$. If there is a sequence $a_n \to a$ such that $q_x(a_n) > 0$, then Condition 4.2 will not be satisfied. The obstacle is encountered only when there is $(x,a) \in S \times A$ such that $q_x(a) = 0$. In this connection, let us mention, when considering an α-discounted CTMDP problem, one can equivalently reformulate it as a total undiscounted one by adding additional transition rate of $\alpha > 0$ at each state to

a cemetery point. The resulting model is with the extended state space, but with a transition rate separated from zero.

One referee suggested the present form of Condition 4.1(b). The next simple example demonstrates that under Condition 4.1(c), the present Condition 4.1(b) is strictly weaker than Condition 4.2. (It is obvious that Condition 4.1(b) is weaker than Condition 4.2 under Condition 4.1(c).)

Example 4.1. Let $S = [0, \infty)$ and $A = [0, 1]$. When $S \times A \ni (x, a) \neq (0, 0)$, $\tilde{q}(dy|x, a) = (x + a)\delta_{x+a+\ln(1+x)}(dy)$; and $q_0(0) = 0$. Then Condition 4.1(b) is satisfied. Note, the state 0 is not a cemetery, there is an absorbing action ($a = 0$) for it, but there are also actions which make the state not absorbing. In view of the presence of this obstacle, it is clear that Condition 4.2 is not satisfied.

The motivation for imposing Condition 4.2 in the original version of the paper is as follows. Under Condition 4.1(a,c,d) and Condition 4.2, the DTMPD model $\{S_\infty, A, p\}$ with the cost functions given by $\frac{c_i(x,a)}{q_x(a)}$ $(i = 0, 1, 2, \ldots, N)$ is semicontinuous in the sense of for example, [5]. Then by Dufour, Horiguchi and Piunovskiy [10], if there is a feasible policy with a finite value for the DTMDP problem (3.3), so is there a stationary optimal one. On the other hand, this statement still holds under Condition 4.1; we formulate this observation in the next lemma.

Lemma 4.2. *Suppose Condition 4.1 is satisfied, and there is a feasible policy with a finite value for the DTMDP problem* (3.3). *Then there is a stationary optimal policy for problem* (3.3).

Before we prove this lemma, let us introduce the following sets

$$\begin{aligned}
S_1 &:= \left\{ x \in S : \inf_{a\in A} q_x(a) = 0, \inf_{a\in A} \left(q_x(a) + \sum_{i=0}^{N} c_i(x, a) \right) > 0 \right\}, \\
\hat{S}_1 &:= \left\{ x \in S_1 : \sup_{a\in A} q_x(a) = 0 \right\}, \\
S_2 &:= \left\{ x \in S : \inf_{a\in A} \left(q_x(a) + \sum_{i=0}^{N} c_i(x, a) \right) = 0 \right\}, \\
S_3 &:= \left\{ x \in S : \inf_{a\in A} q_x(a) > 0 \right\}.
\end{aligned} \tag{4.2}$$

Under Condition 4.1, the above four sets are all measurable, by Proposition 7.32 of [5] and Lemma 4.1. Furthermore, S_1, S_2 and S_3 are disjoint and satisfy

$$S = S_1 \cup S_2 \cup S_3.$$

Let us also denote for each $x \in S$,

$$B(x) := \{ a \in A : q_x(a) = 0 \}, \tag{4.3}$$

which is compact under Condition 4.1.

Proof of Lemma 4.2. Consider the DTMDP model $\{S, A, \tilde{p}, \gamma\}$, where the transition probability $\tilde{p}$ is defined by (4.1). The strategic measure under a policy σ for this DTMDP model is denoted by $\tilde{P}_\gamma^\sigma$, with the corresponding expectation being denoted by $\tilde{E}_\gamma^\sigma$. Consider the following problem

$$\begin{aligned} &\tilde{E}_\gamma^\sigma\left[\sum_{n=0}^{\infty}\frac{c_0(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] \to \min_\sigma \\ &\text{s.t.} \quad \tilde{E}_\gamma^\sigma\left[\sum_{n=0}^{\infty}\frac{c_j(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] \le d_j, \qquad \forall j = 1, 2, \ldots, N. \end{aligned} \tag{4.4}$$

Note that under the same policy, the only difference between the two DTMDP models $\{S_\infty, A, p, \gamma\}$ and $\{S, A, \tilde{p}, \gamma\}$ lies in the following: if the current state x is in $S_1 \cup S_2$, and some action $a \in B(x)$ is chosen, then the process is killed at x_∞ in the first model, whereas the process is not killed but remains in x with a possibility of incurring further costs in the second model.

Consider the feasible policy for problem (3.3) with a finite value; assume without loss of generality that under this policy, when the current state x is in S_2, only action from $B(x)$ will be selected. This policy is also feasible with a finite value for problem (4.4), because under it, the process almost surely never visits $\hat{S}_1$, and when the current state x is in $S_1 \setminus \hat{S}_1$, almost surely no action from $B(x)$ will be selected, for otherwise, it would contradict the assumption of feasibility and finiteness of the value for problem (3.3) under this policy. (Remember the definition of the set S_1.) Now, the observation in the previous paragraph shows that this policy is also feasible with a finite value for problem (4.4).

Under Condition 4.1, the model $\{S, A, \tilde{p}, \gamma\}$ with the cost functions given by $\frac{c_i(x,a)}{q_x(a)}$ is semi-continuous, and so by Theorem 4.1 of [10], there is a stationary optimal policy for problem (4.4). By an argument similar to the above one, one can easily see that this stationary policy is also optimal for problem (3.3). □

On the other hand, Condition 4.1 is exactly the "Continuity and Compactness Assumption" in [18] specified to the case of pure jump processes. A main statement from [18] will be useful for our investigations in the next subsection, which we choose to formulate now as another consequence of Condition 4.1.

Consider the following optimal control problem for the CTMDP model $\{S, A, q\}$:

$$E_x^\pi\left[\int_0^\infty \int_A \sum_{i=0}^{N} c_i(\xi_t, a)\pi(da|\omega, t)\,dt\right] \to \min_{\pi\in\Pi_{\mathrm{DM}}}, \tag{4.5}$$

the value function of which is denoted as

$$V(x) := \inf_{\pi\in\Pi_{\mathrm{DM}}} E_x^\pi\left[\int_0^\infty \int_A \sum_{i=0}^{N} c_i(\xi_t, a)\pi(da|\omega, t)\,dt\right]. \tag{4.6}$$

Here Π_{DM} stands for the class of deterministic Markov policies for the CTMDP model $\{S, A, q\}$. A policy $\pi^* \in \Pi_{\mathrm{DM}}$ is called optimal for problem (4.5) if

$$E_x^{\pi^*}\left[\int_0^\infty \int_A \sum_{i=0}^N c_i(\xi_t, a)\pi^*(da|\omega, t)\,dt\right] = V(x)$$

for each $x \in S$. The following proposition is borrowed from [18]; see Proposition 5.8 and Theorem 5.9 therein.

Proposition 4.1. *Suppose that Condition 4.1 is satisfied. Then the following assertions hold.*

(a) *The function V is the minimal nonnegative lower semicontinuous solution on S to the following Bellman (optimality) equation*:

$$V(x) = \inf_{a\in A}\left\{\frac{\sum_{i=0}^N c_i(x,a)}{q_x(a)} + \int_S \frac{\tilde{q}(dy|x,a)}{q_x(a)} V(y)\right\} \tag{4.7}$$

for each $x \in S$.

(b) *There is a deterministic stationary optimal policy φ^* for the CTMDP problem (4.5), which can be taken as a measurable mapping from S to A such that*

$$\begin{aligned}&\inf_{a\in A}\left\{\frac{\sum_{i=0}^N c_i(x,a)}{q_x(a)} + \int_S \frac{\tilde{q}(dy|x,a)}{q_x(a)} V(y)\right\}\\ &\quad = \frac{\sum_{i=0}^N c_i(x,\varphi^*(x))}{q_x(\varphi^*(x))} + \int_S \frac{\tilde{q}(dy|x,\varphi^*(x))}{q_x(\varphi^*(x))} V(y), \qquad \forall x \in S.\end{aligned}$$

In fact, each deterministic stationary optimal policy for problem (4.5) φ^* satisfies the above relation.

In fact, the authors of [18] considered the more general piecewise deterministic Markov decision process but in the state space $\mathbb{R}^n$. When specializing to the case of a CTMDP, one can put the more general Borel state space S. Furthermore, the authors of [18] assumed that $V(x) < \infty$ for each $x \in S$; see "Boundedness Assumption" in page 252 therein, which, could be withdrawn when specializing to the CTMDP problem (4.5), as far as the validity of the above proposition is concerned.

The previous discussions basically explain why we could replace Condition 4.2 with Condition 4.1(b).

4.2. Properties of occupancy measure

Definition 4.1. *For each $n = 0, 1, \ldots$, the (undiscounted) occupancy measure of the policy $\pi \in \Pi$ for the CTMDP model $\{S, A, q, \gamma\}$ is a measure $M_\gamma^{n,\pi}$ on $\mathcal{B}(S \times A)$ defined by*

$$M_\gamma^{n,\pi}(\Gamma_S \times \Gamma_A) := E_\gamma^\pi\left[\int_{t_n}^{t_{n+1}} I\{\xi_t \in \Gamma_S\} \int_{\Gamma_A} q_{\xi_t}(a)\pi(da|\omega, t)\,dt\right] \tag{4.8}$$

for each $\Gamma_S \in \mathcal{B}(S)$ *and* $\Gamma_A \in \mathcal{B}(A)$.

Lemma 4.3. *Suppose Condition 4.1 is satisfied. Consider a feasible policy* $\pi = (\pi_n) \in \Pi$ *with a finite value for the CTMDP problem (2.2). Then for each* $n = 0, 1, \ldots,$

$$P_\gamma^\pi(x_n \in S_1 \setminus \hat{S}_1) = E_\gamma^\pi\left[I\{x_n \in S_1 \setminus \hat{S}_1\} P_\gamma^\pi\left(\int_0^\infty \int_{A\setminus B(x_n)} q_{x_n}(a)\pi_{n+1}(da|x_0, \theta_1, \ldots, x_n, s)\, ds = \infty \Big| x_n\right)\right].$$

Proof. It holds that for each $n = 0, 1, \ldots,$

$$\begin{aligned}
\infty &> E_\gamma^\pi\left[\int_{t_n}^{t_{n+1}} \int_{(A\setminus B(x_n))\cup B(x_n)} \sum_{i=0}^N c_i(x_n, a)\pi_{n+1}(da|x_0, \ldots, x_n, t)\, dt\right] \\
&\geq E_\gamma^\pi\left[\int_{t_n}^{t_{n+1}} I\{x_n \in S_1 \setminus \hat{S}_1\} \min_{a\in B(x_n)}\left\{\sum_{i=0}^N c_i(x_n, a)\right\} ds\right] \\
&= E_\gamma^\pi\left[I\{x_n \in S_1 \setminus \hat{S}_1\} \min_{a\in B(x_n)}\left\{\sum_{i=0}^N c_i(x_n, a)\right\}\right. \\
&\quad \left.\times \int_0^\infty e^{-\int_0^t \int_A q_{x_n}(a)\pi_{n+1}(da|x_0,\ldots,x_n,s)\, ds}\, dt\right].
\end{aligned} \tag{4.9}$$

If the statement of the lemma does not hold, then there is some $n = 0, 1, \ldots$ such that

$$P_\gamma^\pi\left(x_n \in S_1 \setminus \hat{S}_1, \int_0^\infty \int_A q_{x_n}(a)\pi_{n+1}(da|x_0, \theta_1, \ldots, x_n, s)\, ds < \infty\right) > 0,$$

and thus

$$P_\gamma^\pi\left(x_n \in S_1 \setminus \hat{S}_1, \int_0^\infty e^{-\int_0^t \int_A q_{x_n}(a)\pi_{n+1}(da|x_0,\theta_1,\ldots,x_n,s)\, ds}\, dt = \infty\right) > 0.$$

This implies

$$E_\gamma^\pi\left[I\{x_n \in S_1 \setminus \hat{S}_1\} \min_{a\in B(x_n)}\left\{\sum_{i=0}^N c_i(x_n, a)\right\} \int_0^\infty e^{-\int_0^t \int_A q_{x_n}(a)\pi_{n+1}(da|x_0,\ldots,x_n,s)\, ds}\, dt\right] = \infty,$$

where the last equality follows from the fact that $\min_{a\in B(x)}\{\sum_{i=0}^N c_i(x, a)\} > 0$ for each $x \in S_1$. This contradicts (4.9). □

Definition 4.2. *For each fixed $n = 0, 1, \ldots,$ and policy $\pi = (\pi_n) \in \Pi$, we define a measure $m^{\pi}_{\gamma,n}(dx \times da)$ for the CTMDP model $\{S, A, q, \gamma\}$ on $\mathcal{B}(S \times A)$ by*

$$m^{\pi}_{\gamma,n}(\Gamma_S \times \Gamma_A) := E^{\pi}_{\gamma}\left[\int_{t_n}^{t_{n+1}} I\{x_n \in \Gamma_S\}\pi_{n+1}(\Gamma_A|x_0, \theta_1, \ldots, x_n, t - t_n)\, dt\right] \tag{4.10}$$

for each $\Gamma_S \in \mathcal{B}(S)$ and $\Gamma_A \in \mathcal{B}(A)$.

Evidently, for each $n = 0, 1, \ldots,$ and $\Gamma_S \in \mathcal{B}(S)$, $m^{\pi}_{\gamma,n}(\Gamma_S \times A) > 0$ if and only if $P^{\pi}_{\gamma}(x_n \in \Gamma_S) > 0$.

Lemma 4.4. *Suppose Condition 4.1 is satisfied. Consider a feasible policy $\pi \in \Pi$ with a finite value for the CTMDP problem (2.2). Then it holds that*

$$\int_{S\times A} \tilde{q}(\hat{S}_1|x, a) m^{\pi}_{\gamma,n}(dx \times da) = 0,$$

and

$$m^{\pi}_{\gamma,n}(\hat{S}_1 \times A) = 0 \tag{4.11}$$

for each $n = 0, 1, \ldots$. In particular,

$$\gamma(\hat{S}_1) = 0.$$

Proof. Suppose for contradiction that $m^{\pi}_{\gamma,n}(\hat{S}_1 \times A) > 0$ for some n. Then similarly to the proof of Lemma 4.3, one can establish the following contradiction;

$$\begin{aligned}\infty &> E^{\pi}_{\gamma}\left[\int_{t_n}^{t_{n+1}} \int_A \sum_{i=1}^{N} c_i(x_n, a)\pi(da|\omega, t)\, dt\right] \\ &\geq E^{\pi}_{\gamma}\left[\int_{t_n}^{t_{n+1}} \min_{a\in A}\left\{\sum_{i=1}^{N} c_i(x_n, a)\right\} I\{x_n \in \hat{S}_1\}\, dt\right] = \infty.\end{aligned}$$

As a result, $m^{\pi}_{\gamma,n}(\hat{S}_1 \times A) = 0$ for each $n = 0, 1, \ldots$. In particular, $m^{\pi}_{\gamma,0}(\hat{S}_1 \times A) = 0$, and thus $P^{\pi}_{\gamma}(x_0 \in \hat{S}_1) = \gamma(\hat{S}_1) = 0$. It remains to prove $\int_{S\times A} \tilde{q}(\hat{S}_1|x, a) m^{\pi}_{\gamma,n}(dx \times da) = 0$ for each $n = 0, 1, \ldots$. If this is not true, then it follows from the definition of $m^{\pi}_{\gamma,n}(dx \times da)$ that for some $n = 0, 1, \ldots,$

$$\begin{aligned}0 &< E^{\pi}_{\gamma}\left[\int_0^{\theta_{n+1}} \int_A \tilde{q}(\hat{S}_1|x_n, a)\pi_{n+1}(da|x_0, \theta_1, \ldots, x_n, t)\, dt\right] \\ &= E^{\pi}_{\gamma}\left[\int_0^{\infty} \int_A \tilde{q}(\hat{S}_1|x_n, a)\pi_{n+1}(da|x_0, \theta_1, \ldots, x_n, t)\right. \\ &\qquad \left. \times e^{-\int_0^t \int_A q_{x_n}(a)\pi_{n+1}(da|x_0,\theta_1,\ldots,x_n,s)\, ds}\, dt\right],\end{aligned}$$

which implies that $P_\gamma^\pi(x_{n+1} \in \hat{S}_1) > 0$ by the construction of the CTMDP; see (2) of [29]. This leads to the contradiction against the fact that $m_{\gamma,n+1}^\pi(\hat{S}_1 \times A) = 0$ as established earlier. □

Definition 4.3. *Let f^* be a fixed measurable mapping from S to A such that*

$$0 = \inf_{a \in A} \left\{ \sum_{i=0}^{N} c_i(x,a) + q_x(a) \right\} = \sum_{i=0}^{N} c_i\big(x, f^*(x)\big) + q_x\big(f^*(x)\big) \tag{4.12}$$

for each $x \in S_2$ whenever S_2 is nonempty.

The existence of such a mapping is guaranteed by Proposition 7.33 of Bertsekas and Shreve [5] under Condition 4.1.

Theorem 4.1. *Suppose Condition 4.1 is satisfied. Consider a feasible policy $\pi = (\pi_n) \in \Pi$ with a finite value for the CTMDP problem (2.2) such that*

$$\pi_{n+1}(da|x_0, \theta_1, \dots, x_n, s) = \delta_{f^*(x_n)}(da) \tag{4.13}$$

whenever $x_n \in S_2$. Then there is a Markov policy σ for the DTMDP $\{S_\infty, A, p, \gamma\}$ such that for each $n = 0, 1, \dots,$

$$\sigma_{n+1}(da|x) = \delta_{f^*(x)}(da) \tag{4.14}$$

for each $x \in S_2$ (if $S_2 \neq \varnothing$), and

$$M_\gamma^{n,\pi}(\Gamma_S \times \Gamma_A) = \mathbf{P}_\gamma^\sigma(X_n \in \Gamma_S, A_{n+1} \in \Gamma_A), \qquad \forall \Gamma_S \in \mathcal{B}(S \setminus S_2), \Gamma_A \in \mathcal{B}(A). \tag{4.15}$$

Proof. For each $\Gamma_S \in \mathcal{B}(S)$,

$$\begin{aligned}
&M_\gamma^{n,\pi}(\Gamma_S \times A) \\
&\quad = E_\gamma^\pi\left[E_\gamma^\pi\left[\int_{t_n}^{t_{n+1}} I\{\xi_t \in \Gamma_S\} \int_A q_{\xi_t}(a)\pi(da|\omega,t)\,dt \,|\, x_0, \theta_1, \dots, x_n \right] \right] \\
&\quad = E_\gamma^\pi\left[I\{x_n \in \Gamma_S\} E_\gamma^\pi\left[\int_0^{\theta_{n+1}} \int_A q_{x_n}(a)\pi_{n+1}(da|x_0, \theta_1, \dots, x_n, t)\,dt \,|\, x_0, \theta_1, \dots, x_n \right] \right] \\
&\quad = E_\gamma^\pi\left[I\{x_n \in \Gamma_S\} \int_0^\infty \int_A q_{x_n}(a)\pi_{n+1}(da|x_0, \theta_1, \dots, x_n, t) \right. \\
&\qquad \left. \times e^{-\int_0^t \int_A q_{x_n}(a)\pi_{n+1}(da|x_0,\theta_1,\dots,x_n,s)\,ds}\,dt \right] \\
&\quad = E_\gamma^\pi\left[I\{x_n \in \Gamma_S\}\left(1 - e^{-\int_0^\infty \int_A q_{x_n}(a)\pi_{n+1}(da|x_0,\theta_1,\dots,x_n,s)\,ds}\right)\right] \le 1.
\end{aligned} \tag{4.16}$$

Then for each $n = 0, 1, \ldots,$ one can refer to Corollary 7.27.2 of [5] or Proposition D.8 of [23] for the existence of a stochastic kernel $\sigma_{n+1}(da|x)$ such that

$$M_\gamma^{n,\pi}(dx \times da) = M_\gamma^{n,\pi}(dx \times A)\sigma_{n+1}(da|x) \tag{4.17}$$

on $\mathcal{B}(S \times A)$, and (4.14) holds, where the last assertion is true because $M_\gamma^{n,\pi}(S_2 \times A) = 0$ by (4.13), (4.12) and (4.8). Let $\sigma = (\sigma_n)$ be the Markov policy for the DTMDP model $\{S_\infty, A, p, \gamma\}$ defined by this sequence of stochastic kernels.

Consider the case of $n = 0$. Then for each $\Gamma_S \in \mathcal{B}(S \setminus S_2)$,

$$\begin{aligned} M_\gamma^{0,\pi}(\Gamma_S \times A) &= E_\gamma^\pi\big[I\{x_0 \in \Gamma_S\}\big(1 - e^{-\int_0^\infty \int_A q_{x_0}(a)\pi_1(da|x_0,s)\,ds}\big)\big] \\ &= \gamma(\Gamma_S) = \mathbf{P}_\gamma^\sigma(X_0 \in \Gamma_S), \end{aligned} \tag{4.18}$$

where the first equality is by (4.16), the second equality follows from Lemma 4.4 in case $\Gamma_S \subseteq \hat{S}_1$, from Lemma 4.3 in case $\Gamma_S \subseteq S_1 \setminus \hat{S}_1$, and from (4.2) in case $\Gamma_S \subseteq S_3$. Consequently, for each $\Gamma_S \in \mathcal{B}(S \setminus S_2)$ and $\Gamma_A \in \mathcal{B}(A)$,

$$\begin{aligned} M_\gamma^{0,\pi}(\Gamma_S \times \Gamma_A) &= \int_{\Gamma_S} M_\gamma^{0,\pi}(dx \times A)\sigma_1(\Gamma_A|x) = \int_{\Gamma_S} \mathbf{P}_\gamma^\sigma(X_0 \in dx)\sigma_1(\Gamma_A|x) \\ &= \mathbf{P}_\gamma^\sigma(X_0 \in \Gamma_S, A_1 \in \Gamma_A), \end{aligned} \tag{4.19}$$

where the first equality is by (4.17).

Suppose that (4.15) holds for all $n \le k$. Consider the case of $n = k + 1$ as follows.

Note that for each $n = 0, 1, \ldots,$

$$M_\gamma^{n,\pi}(\hat{S}_1 \times A) = 0 = M_\gamma^{n,\pi}(S_2 \times A),$$

where the first equality is by Lemma 4.4, and the second equality is by (4.12) and (4.13). Now

$$\int_{S_2 \times A} \frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)} \mathbf{P}_\gamma^\sigma(X_k \in dy, A_{k+1} \in da) = 0 = \int_{S_2 \times A} \frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)} M_\gamma^{k,\pi}(dy \times da).$$

Consequently, for each $\Gamma_S \in \mathcal{B}(S \setminus S_2)$,

$$\begin{aligned} &\mathbf{P}_\gamma^\sigma(X_{k+1} \in \Gamma_S) \\ &= \int_{S \times A} \frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)} \mathbf{P}_\gamma^\sigma(X_k \in dy, A_{k+1} \in da) \\ &= \int_{(S \setminus S_2) \times A} \frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)} \mathbf{P}_\gamma^\sigma(X_k \in dy, A_{k+1} \in da) \\ &\quad + \int_{S_2 \times A} \frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)} \mathbf{P}_\gamma^\sigma(X_k \in dy, A_{k+1} \in da) \\ &= \int_{(S \setminus S_2) \times A} \frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)} M_\gamma^{k,\pi}(dy \times da) \end{aligned} \tag{4.20}$$

$$+\int_{S_2\times A}\frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)}M_\gamma^{k,\pi}(dy\times da)$$
$$=\int_{S\times A}\frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)}M_\gamma^{k,\pi}(dy\times da).$$

On the other hand, for each $\Gamma_S\in\mathcal{B}(S\setminus S_2)$,

$$\begin{aligned}
&\int_{S\times A}\frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)}M_\gamma^{k,\pi}(dy\times da)\\
&\quad=\int_{S\times A}\frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)}q_y(a)m_{\gamma,k}^{\pi}(dy\times da)\\
&\quad=\int_{S\times A}\tilde{q}(\Gamma_S|y,a)m_{\gamma,k}^{\pi}(dy\times da)\\
&\quad=E_\gamma^\pi\left[\int_{t_k}^{t_{k+1}}\int_A\tilde{q}(\Gamma_S|\xi_t,a)\pi(da|\omega,t)\,dt\right]\\
&\quad=E_\gamma^\pi\left[\int_0^{\theta_{k+1}}\int_A\tilde{q}(\Gamma_S|x_k,a)\pi_{k+1}(da|x_0,\theta_1,\dots,x_k,t)\,dt\right]\\
&\quad=E_\gamma^\pi\left[\int_0^{\infty}\int_A\tilde{q}(\Gamma_S|x_k,a)\pi_{k+1}(da|x_0,\theta_1,\dots,x_k,t)e^{-\int_0^t\int_A q_{x_k}(a)\pi_{k+1}(da|x_0,\dots,x_k,s)\,ds}\,dt\right]\\
&\quad=E_\gamma^\pi\left[\frac{\int_A\tilde{q}(\Gamma_S|x_k,a)\pi(da|\omega,t_{k+1})}{\int_A q_{x_k}(a)\pi(da|\omega,t_{k+1})}\right],
\end{aligned}\tag{4.21}$$

where the first and the third equalities are by (4.10), whereas the second equality follows from the fact that if $q_y(a)=0$, then

$$\tilde{q}(\Gamma_S|y,a)=0=\frac{\tilde{q}(\Gamma_S|y,a)}{q_y(a)}q_y(a)$$

keeping in mind $\frac{0}{0}=0$; the similar reasoning justifies the last equality, too. This together with (4.20) shows

$$\mathbf{P}_\gamma^\sigma(X_{k+1}\in\Gamma_S)=E_\gamma^\pi\left[\frac{\int_A\tilde{q}(\Gamma_S|x_k,a)\pi(da|\omega,t_{k+1})}{\int_A q_{x_k}(a)\pi(da|\omega,t_{k+1})}\right]\tag{4.22}$$

for each $\Gamma_S\in\mathcal{B}(S\setminus S_2)$.

Now it holds that

$$\begin{aligned}
\mathbf{P}_\gamma^\sigma(X_{k+1}\in\hat{S}_1)&=\int_{S\times A}\tilde{q}(\hat{S}_1|x,a)m_k^\pi(dx\times da)\\
&=0=M_\gamma^{k+1,\pi}(\hat{S}_1\times A),
\end{aligned}\tag{4.23}$$

where the first equality is by (4.21), whereas the second and the last equalities are by Lemma 4.4.

One can see that for each $\Gamma_S \in \mathcal{B}(S \setminus (S_2 \cup \hat{S}_1))$,

$$\begin{aligned} M_\gamma^{k+1,\pi}(\Gamma_S \times A) &= E_\gamma^\pi\big[I\{x_{k+1} \in \Gamma_S\}\big] \\ &= E_\gamma^\pi\big[E_\gamma^\pi\big[I\{x_{k+1} \in \Gamma_S\}|x_0, \theta_1, \dots, x_k, \theta_{k+1}\big]\big] \\ &= E_\gamma^\pi\left[\frac{\int_A \tilde{q}(\Gamma_S|x_k, a)\pi(da|\omega, t_{k+1})}{\int_A q_{x_k}(a)\pi(da|\omega, t_{k+1})}\right] = \mathbf{P}_\gamma^\sigma(X_{k+1} \in \Gamma_S), \end{aligned}$$

where the first equality is by the last equality of (4.16) keeping in mind Lemma 4.3 and (4.2), and the last equality is by (4.22). This and (4.23) justify that

$$M_\gamma^{k+1,\pi}(\Gamma_S \times A) = \mathbf{P}_\gamma^\sigma(X_{k+1} \in \Gamma_S)$$

for each $\Gamma_S \in \mathcal{B}(S \setminus S_2)$. Now we see

$$\begin{aligned} M_\gamma^{k+1,\pi}(\Gamma_S \times \Gamma_A) &= \int_{\Gamma_S} M_\gamma^{k+1,\pi}(dx \times A)\sigma_{k+2}(\Gamma_A|x) \\ &= \int_{\Gamma_S} \mathbf{P}_\gamma^\sigma(X_{k+1} \in dx)\sigma_{k+2}(\Gamma_A|x) = \mathbf{P}_\gamma^\sigma(X_{k+1} \in \Gamma_S, A_{k+2} \in \Gamma_A) \end{aligned}$$

for each $\Gamma_S \in \mathcal{B}(S \setminus S_2)$ and $\Gamma_A \in \mathcal{B}(A)$. The statement of the theorem is thus proved by induction. □

5. Occupation measure

The objective of this section is to show that restricted on a measurable subset $\zeta \subseteq S$, the measure $\eta_\gamma^\pi(dx \times A)$ to be defined below is σ-finite; see Theorem 5.1 below, where ζ is defined by (5.10), whereas the set ζ^c is easy to deal with. We do this by using the technique developed in [10] by Dufour, Horiguchi and Piunovskiy, who dealt with DTMDP problems.

Definition 5.1. *For each policy $\pi \in \Pi$ for the CTMDP model (S, A, q, γ), its (undiscounted) occupation measure η_γ^π is the measure on $\mathcal{B}(S \times A)$ given by*

$$\eta_\gamma^\pi(\Gamma_S \times \Gamma_A) := E_\gamma^\pi\left[\int_0^\infty I\{\xi_t \in \Gamma_S\}\pi(\Gamma_A|\omega, t)\,dt\right]$$

for each $\Gamma_S \in \mathcal{B}(S)$ and $\Gamma_A \in \mathcal{B}(A)$.

Lemma 5.1. *For each policy π for the CTMDP model, its (undiscounted) occupation measure η_γ^π satisfies the following relation*:

$$\int_{\Gamma \times A} q_x(a)\eta_\gamma^\pi(dx \times da) + Z^\pi(\Gamma) = \gamma(\Gamma) + \int_{S \times A} \tilde{q}(\Gamma|y, a)\eta_\gamma^\pi(dy \times da) \tag{5.1}$$

for each $\Gamma \in \mathcal{B}(S)$, *where* $Z^{\pi}(\Gamma) \in [0,1]$.

Proof. For each $\alpha > 0$, consider the measure on $\mathcal{B}(S \times A)$

$$\eta_{\gamma}^{\alpha,\pi}(dx \times da) = E_{\gamma}^{\pi}\left[\int_0^{\infty} e^{-\alpha t} I\{\xi_t \in dx\}\pi(da|\omega,t)\,dt\right], \tag{5.2}$$

which is the (α-discounted) occupation measure of the policy π for the CTMDP model $\{S, A, q, \gamma\}$. It follows from the definition that for each $\pi \in \Pi$,

$$\big(\alpha + q_x(a)\big)\eta_{\gamma}^{\alpha,\pi}(dx \times da) = \sum_{n=0}^{\infty} M_{\gamma,\alpha}^{n,\pi}(dx \times da).$$

By Proposition 3.1, there is some policy σ for the DTMDP model $\{S_{\infty}, A, p_{\alpha}, \gamma\}$ satisfying (3.2) on $\mathcal{B}(S \times A)$. Note that $\sum_{n=0}^{\infty} P_{\gamma}^{\alpha,\sigma}(X_n \in dx, A_{n+1} \in da)$, the right-hand side of (3.2), is the undiscounted occupation measure for the DTMDP model $\{S_{\infty}, A, p_{\alpha}, \gamma\}$ restricted to $S \times A$, so that, by a well known and easy-to-see fact from the theory of DTMDPs, for each $\Gamma \in \mathcal{B}(S)$,

$$\sum_{n=0}^{\infty} P_{\gamma}^{\alpha,\sigma}(X_n \in \Gamma) = \gamma(\Gamma) + \int_{S\times A} \frac{\tilde{q}(\Gamma|x,a)}{\alpha + q_x(a)} \sum_{n=0}^{\infty} P_{\gamma}^{\alpha,\sigma}(X_n \in dx, A_{n+1} \in da).$$

By (3.2), the above can be written as

$$\int_{\Gamma\times A} \big(q_x(a) + \alpha\big)\eta_{\gamma}^{\alpha,\pi}(dx \times da) = \gamma(\Gamma) + \int_{S\times A} \tilde{q}(\Gamma|x,a)\eta_{\gamma}^{\alpha,\pi}(dx \times da). \tag{5.3}$$

Keeping in mind

$$\int_{\Gamma\times A} \alpha\eta_{\gamma}^{\alpha,\pi}(dx \times da) = E_{\gamma}^{\pi}\left[\int_0^{\infty} \alpha e^{-\alpha t} I\{\xi_t \in \Gamma\}\,dt\right] \in [0,1]$$

for each $\alpha \in (0,\infty)$, one can legitimately take the upper limit as $0 < \alpha \downarrow 0$ on the both sides of the above equality to see that $\eta_{\gamma}^{\pi}(dx \times da)$ satisfies that for each $\Gamma \in \mathcal{B}(S)$

$$\begin{aligned}&\int_{\Gamma\times A} q_x(a)\eta_{\gamma}^{\pi}(dx \times da) + \limsup_{0<\alpha\downarrow 0} \int_{\Gamma\times A} \alpha\eta_{\gamma}^{\alpha,\pi}(dx \times da)\\ &\quad = \gamma(\Gamma) + \int_{S\times A} \tilde{q}(\Gamma|y,a)\eta_{\gamma}^{\pi}(dy \times da),\end{aligned}$$

where we have used the fact that $\eta_{\gamma}^{\alpha,\pi}(dx \times da) \uparrow \eta_{\gamma}^{\pi}(dx \times da)$ setwise as $\alpha \downarrow 0$, and the monotone convergence theorem; see Theorem 2.1 of [24]. By putting $Z^{\pi}(\Gamma) = \limsup_{0<\alpha\downarrow 0} \int_{\Gamma\times A} \alpha\eta_{\gamma}^{\alpha,\pi}(dx \times da) \in [0,1]$, we see that the statement of the lemma holds. □

Remark 5.1. The relation (5.3) was established under the extra conditions imposed on the growth of the transition rates $q(dy|x,a)$ in [30]. The relation (5.1) was established for certain subsets

$\Gamma \in \mathcal{B}(S)$ in [22], where the authors imposed extra conditions and considered the absorbing models, so that the term $Z^\pi(\Gamma)$ vanishes for all the "transient" subsets Γ.

Lemma 5.2. *Suppose Condition 4.1(a) is satisfied, and let an extended real-valued lower semi-continuous function g on $S \times A$ be fixed. Then the following assertions hold.*

(a) *For each $\varepsilon \in \mathbb{R}$, it holds that the set $\{x \in S : \forall a \in A, g(x,a) > \varepsilon\}$ is open in S.*
(b) $\{x \in S : \forall a \in A, g(x,a) > 0\} = \bigcup_{l=1}^{\infty} \{x \in S : \forall a \in A, g(x,a) > \frac{1}{l}\}$.

Proof. See Lemmas 3.1 and 3.2 in [10]. □

Lemma 5.3. *Let some feasible policy π for problem (2.2) with a finite value be fixed. Suppose that Condition 4.1 is satisfied, and that $\{B_j, j = 1, 2, \ldots\} \subseteq \mathcal{B}(S)$ is an increasing sequence of open sets satisfying*

$$\int_{B_j \times A} q_x(a) \eta_\gamma^\pi(dx \times da) < \infty. \tag{5.4}$$

Then the following assertions hold.

(a) *There exists a sequence of open sets $\{E_j, j = 1, 2, \ldots\} \subseteq \mathcal{B}(S)$ such that*

$$Y := \left\{ x \in S : \forall a \in A, \tilde{q}\left(\bigcup_j B_j \middle| x, a \right) + \sum_{i=0}^{N} c_i(x,a) > 0 \right\} = \bigcup_j E_j,$$

and for all $j = 1, 2, \ldots, \int_{E_j \times A} q_x(a) \eta_\gamma^\pi(dx \times da) < \infty$.

(b) *There exists a sequence of open sets $\{\tilde{E}_j, j = 1, 2, \ldots\} \subseteq \mathcal{B}(S)$ such that $Y = \bigcup_j \tilde{E}_j$, and for all $j = 1, 2, \ldots, \eta_\gamma^\pi(\tilde{E}_j \times A) < \infty$.*

Proof. (a) Define for each $l = 1, 2, \ldots$ and $j = 1, 2, \ldots,$

$$B_j^{(l)} := \left\{ (x,a) \in S \times A : \frac{\tilde{q}(B_j|x,a) + \sum_{i=0}^{N} c_i(x,a)}{q_x(a)} > \frac{1}{l} \right\}$$

and

$$C_j^{(l)} := \left\{ x \in S : \forall a \in A, \frac{\tilde{q}(B_j|x,a) + \sum_{i=0}^{N} c_i(x,a)}{q_x(a)} > \frac{1}{l} \right\}.$$

From (5.1), we see that (5.4) implies

$$\int_{S \times A} \tilde{q}(B_j|x,a) \eta_\gamma^\pi(dx \times da) < \infty. \tag{5.5}$$

Let $N(q) := \{(x,a) \in S \times A : q_x(a) = 0\}$. Then for each $j, l = 1, 2, \ldots,$

$$\begin{aligned}
&\int_{B_j^{(l)}} q_x(a)\eta_\gamma^\pi(dx \times da) \\
&= \int_{B_j^{(l)} \cap (N(q)^c)} q_x(a)\eta_\gamma^\pi(dx \times da) \\
&\le \int_{B_j^{(l)} \cap (N(q)^c)} q_x(a) \frac{\tilde{q}(B_j|x,a) + \sum_{i=0}^N c_i(x,a)}{q_x(a)} l\eta_\gamma^\pi(dx \times da) \\
&\le l\left(\int_{S\times A} \tilde{q}(B_j|x,a)\eta_\gamma^\pi(dx \times da) + \int_{S\times A} \sum_{i=0}^N c_i(x,a)\eta_\gamma^\pi(dx \times da) \right) < \infty,
\end{aligned}$$

where the last inequality follows from (5.5) and the assumption of the policy π being feasible with a finite value. Since $C_j^{(l)} \times A \subseteq B_j^{(l)}$, it follows that for each $l, j = 1, 2, \ldots,$

$$\int_{C_j^{(l)} \times A} q_x(a)\eta_\gamma^\pi(dx \times da) \le \int_{B_j^{(l)}} q_x(a)\eta_\gamma^\pi(dx \times da) < \infty. \tag{5.6}$$

Since B_j is open in S for each $j = 1, 2, \ldots, \tilde{q}(B_j|x,a) + \sum_{i=0}^N c_i(x,a) \ge 0$ is lower semicontinuous in $(x,a) \in S \times A$ according to Lemma 4.1(b). By Lemma 4.1(a), $\frac{\tilde{q}(B_j|x,a)+\sum_{i=0}^N c_i(x,a)}{q_x(a)}$ is lower semicontinuous in $(x,a) \in S \times A$. Now referring to Lemma 5.2(a), we see that $C_j^{(l)}$ is open for each $j, l = 1, 2, \ldots$.

Next, let us show

$$\bigcup_j \bigcup_l C_j^{(l)} = Y = \left\{ x \in S : \forall a \in A, \tilde{q}\left(\bigcup_j B_j \middle| x, a \right) + \sum_{i=0}^N c_i(x,a) > 0 \right\} \tag{5.7}$$

as follows. By Lemma 5.2(b), for each $j = 1, 2, \ldots,$

$$\bigcup_l C_j^{(l)} = \left\{ x \in S : \forall a \in A, \frac{\tilde{q}(B_j|x,a) + \sum_{i=0}^N c_i(x,a)}{q_x(a)} > 0 \right\},$$

so that

$$\bigcup_j \bigcup_l C_j^{(l)} \subseteq Y.$$

For the opposite direction of the above relation, we argue as follows. Let some $y \in Y$ be arbitrarily fixed. Since $\frac{\tilde{q}(B_j|x,a)+\sum_{i=0}^N c_i(x,a)}{q_x(a)}$ is lower semicontinuous in $(x,a) \in S \times A$ as explained earlier, and is increasing in $j = 1, 2, \ldots$ keeping in mind that $\{B_j\}$ is an increasing sequence, one

can refer to Proposition 10.1 of [35] (see also Appendix A of [4]) for the following interchange of the order of infimum and limit:

$$\lim_{j\to\infty}\inf_{a\in A}\left\{\frac{\tilde{q}(B_j|y,a)+\sum_{i=0}^{N}c_i(y,a)}{q_y(a)}\right\}=\inf_{a\in A}\lim_{j\to\infty}\left\{\frac{\tilde{q}(B_j|y,a)+\sum_{i=0}^{N}c_i(y,a)}{q_y(a)}\right\}$$
$$=\inf_{a\in A}\left\{\frac{\tilde{q}(\bigcup_j B_j|y,a)+\sum_{i=0}^{N}c_i(y,a)}{q_y(a)}\right\}>0,$$

where the last inequality follows from the fact that $y\in Y$. This implies the existence of some $j=1,2,\ldots$ such that for each $a\in A$, it holds that $\tilde{q}(B_j|y,a)+\sum_{i=0}^{N}c_i(y,a)>0$, i.e., $y\in\bigcup_l C_j^{(l)}\subseteq\bigcup_j\bigcup_l C_j^{(l)}$. Since $y\in Y$ is arbitrarily fixed, this verifies

$$Y\subseteq\bigcup_j\bigcup_l C_j^{(l)}.$$

Hence, (5.7) holds, which in combination with (5.6), proves the statement; remember that $C_j^{(l)}$ is open for each $j,l=1,2,\ldots$.

(b) The proof of this part is similar to the one of part (a). Instead of $B_j^{(l)}$ and $C_j^{(l)}$, one should now introduce for each $j,l=1,2,\ldots,$

$$\tilde{B}_j^{(l)}:=\left\{(x,a)\in S\times A:\tilde{q}(B_j|x,a)+\sum_{i=0}^{N}c_i(x,a)>\frac{1}{l}\right\}$$

and

$$\tilde{C}_j^{(l)}:=\left\{x\in S:\forall a\in A,\tilde{q}(B_j|x,a)+\sum_{i=0}^{N}c_i(x,a)>\frac{1}{l}\right\},$$

so that

$$\eta_\gamma^\pi\big(\tilde{B}_j^{(l)}\big)\le\int_{\tilde{B}_j^{(l)}}l\left(\tilde{q}(B_j|y,a)+\sum_{i=0}^{N}c_i(y,a)\right)\eta_\gamma^\pi(dy\times da)<\infty.$$

Consequently, $\eta_\gamma^\pi(\tilde{C}_j^{(l)}\times A)<\infty$ for each $j,l=1,2,\ldots$. It is clear now how to proceed the rest of the reasoning as in the proof of part (a). □

Lemma 5.4. *Suppose Condition 4.1 is satisfied. Let some feasible policy π for problem (2.2) with a finite value be fixed. Consider*

$$W:=\bigcup_{j=1}^{\infty}W_j,$$

where W_j is defined recursively as follows:

$$W_1 := \left\{ x \in S : \forall a \in A, \frac{\sum_{i=0}^{N} c_i(x,a)}{q_x(a)} > 0 \right\};$$

and for each $j = 1, 2, \ldots,$

$$W_{j+1} := \left\{ x \in S : \forall a \in A, \frac{\tilde{q}(\bigcup_{i=1}^{j} W_i | x, a) + \sum_{i=0}^{N} c_i(x,a)}{q_x(a)} > 0 \right\}.$$

Then for each $j = 1, 2, \ldots, W_j$ is open in S, and so is W. Furthermore, $\eta_\gamma^\pi(dx \times A)$, being restricted to $W \in \mathcal{B}(S)$, is a σ-finite measure on $\mathcal{B}(W)$; in other words, $\eta_\gamma^\pi(dx \times A)$ is σ-finite on W.

Proof. First of all, let us show by induction that for each $m = 1, 2, \ldots, W_m$ is open, and there exists a sequence of open sets $\{E_j^{(m)}\} \subseteq \mathcal{B}(S)$ such that

$$W_m = \bigcup_j E_j^{(m)} \tag{5.8}$$

and

$$\int_{E_j^{(m)} \times A} q_x(a) \eta_\gamma^\pi(dx \times da) < \infty. \tag{5.9}$$

By Lemma 5.2(b), $W_1 = \bigcup_j E_j^{(1)}$, where for each $j = 1, 2, \ldots,$

$$E_j^{(1)} = \left\{ x \in S : \forall a \in A, \frac{\sum_{i=0}^{N} c_i(x,a)}{q_x(a)} > \frac{1}{j} \right\}$$

is open because $\frac{\sum_{i=0}^{N} c_i(x,a)}{q_x(a)}$ is lower semicontinuous in $(x, a) \in S \times A$, and Lemma 5.2(a). Therefore, W_1 is open. Moreover,

$$\int_{E_j^{(1)} \times A} q_x(a) \eta_\gamma^\pi(dx \times da) < \infty$$

because the policy π is feasible with a finite value. Note that $\{E_j^{(1)}, j = 1, 2, \ldots\}$ is an increasing sequence of open sets.

Suppose that for each $k \le n$, W_k is open, and there exists a sequence of open sets $\{E_j^{(k)}, j = 1, 2, \ldots\} \subseteq \mathcal{B}(S)$ such that $W_k = \bigcup_j E_j^{(k)}$ and $\int_{E_j^{(k)} \times A} q_x(a) \eta_\gamma^\pi(dx \times da) < \infty$. Then

$$\begin{aligned} W_{n+1} &= \left\{ x \in S : \forall a \in A, \frac{\tilde{q}(\bigcup_{i=0}^{n} W_i | x, a) + \sum_{i=0}^{N} c_i(x,a)}{q_x(a)} > 0 \right\} \\ &= \left\{ x \in S : \forall a \in A, \frac{\tilde{q}(\bigcup_{i=0}^{n} \bigcup_j E_j^{(i)} | x, a) + \sum_{i=0}^{N} c_i(x,a)}{q_x(a)} > 0 \right\}. \end{aligned}$$

Note that each of the sets $E_j^{(i)}, i = 1, 2, \ldots, n, j = 1, 2, \ldots$ is open, and $\int_{E_j^{(i)} \times A} q_x(a)\eta_\gamma^\pi(dx \times da) < \infty$ by the inductive supposition, so that $\bigcup_{i=0}^n \bigcup_j E_j^{(i)}$ can be rewritten as the union of an increasing sequence of open sets in S, each of which is of finite measure with respect to $\int_A q_x(a)\eta_\gamma^\pi(dx \times da)$. Therefore, one can refer to Lemma 5.3(a) for the existence of a sequence of open sets $\{E_j^{(n+1)}, j = 1, 2, \ldots\} \subseteq \mathcal{B}(S)$ satisfying $W_{n+1} = \bigcup_j E_j^{(n+1)}$ and $\int_{E_j^{(n+1)} \times A} q_x(a)\eta_\gamma^\pi(dx \times da) < \infty$. Thus, W_{n+1} is open in S, and the inductive argument is completed.

We now prove the statement of the lemma. Let us rewrite

$$W_1 = \left\{ x \in S : \forall a \in A, \sum_{i=0}^N c_i(x, a) > 0 \right\}.$$

By Lemma 5.2(b),

$$W_1 = \bigcup_{j=1}^\infty \tilde{E}_j^{(1)},$$

where for each $j = 1, 2, \ldots,$

$$\tilde{E}_j^{(1)} = \left\{ x \in S : \forall a \in A, \sum_{i=0}^N c_i(x, a) > \frac{1}{j} \right\},$$

which is open by Lemma 5.2(a), and satisfies

$$\eta_\gamma^\pi\big(\tilde{E}_j^{(1)} \times A\big) < \infty$$

by the fact that the policy π is feasible with a finite value. For each $m = 2, 3, \ldots,$ by what was established in the beginning of this proof, (5.8), (5.9) and Lemma 5.3(b), which is applicable since $\bigcup_j E_j^{(m)}$ can be rewritten as the union of an increasing sequence of open sets each of finite measure with respect to $\int_A q_x(a)\eta_\gamma^\pi(dx \times da)$, there exists a sequence of open sets $\{\tilde{E}_j^{(m)}, j = 1, 2, \ldots\} \subseteq \mathcal{B}(S)$ such that $W_m = \bigcup_j \tilde{E}_j^{(m)}$ and $\eta_\gamma^\pi(\tilde{E}_j^{(m)} \times A) < \infty$ for each $j = 1, 2, \ldots$. It follows that the statement to be proved holds. □

Definition 5.2. *Let us define the set*

$$\zeta := \left\{ x \in S : \inf_{\pi \in \Pi_{\mathrm{DM}}} E_x^\pi \left[\int_0^\infty \int_A \sum_{i=0}^N c_i(\xi_t, a)\pi(da|\omega, t)\, dt \right] > 0 \right\}. \tag{5.10}$$

Here Π_{DM} stands for the class of deterministic Markov policies for the CTMDP model. Under Condition 4.1, one can refer to Proposition 4.1 for that ζ is a measurable (in fact, open) subset of S.

Theorem 5.1. *Suppose Condition 4.1 is satisfied. Let some feasible policy π for problem (2.2) with a finite value be fixed. Then $\eta_\gamma^\pi(dx \times A)$ is σ-finite on ζ.*

Proof. By Lemma 5.4, the statement of this theorem would be proved if we showed

$$\zeta \subseteq W, \tag{5.11}$$

where the set W is defined in the statement of Lemma 5.4. This fact can be proved in the same way as in the proof of Proposition 3.3 of [10]. □

Definition 5.3. *Suppose Condition 4.1 is satisfied. Let us fix a measurable mapping ψ^* from S to A such that whenever $\zeta^c \neq \varnothing$,*

$$E_x^{\psi^*}\left[\int_0^\infty \left(\sum_{i=0}^N c_i\big(\xi_t, \psi^*(\xi_t)\big)\right) dt\right] = 0 \tag{5.12}$$

for each $x \in \zeta^c$; and whenever $S_2 \neq \varnothing$,

$$\psi^*(x) = f^*(x) \tag{5.13}$$

for each $x \in S_2$, where f^* is defined by (4.12). Such a mapping, or say it interchangeably a deterministic stationary policy, ψ^* exists by Proposition 4.1.

Note that it necessarily holds that for each $i = 0, 1, \ldots, N$,

$$c_i\big(x, \psi^*(x)\big) = 0 \tag{5.14}$$

for all $x \in \zeta^c$, whenever $\zeta^c \neq \varnothing$. Evidently,

$$S_2 \subseteq \zeta^c.$$

The next statement is a direct consequence of Theorem 5.1 and Corollary 7.27.2 of [5].

Corollary 5.1. *Suppose Condition 4.1 is satisfied, and consider a feasible policy $\pi \in \Pi$ for problem (2.2) with a finite value. Then there exists a stationary policy $\varphi_\pi \in \Pi$ such that*

$$\eta_\gamma^\pi(dx \times da) = \varphi_\pi(da|x)\eta_\gamma^\pi(dx \times A) \tag{5.15}$$

on $\mathcal{B}(\zeta \times A)$, and

$$\varphi_\pi(da|x) = \delta_{\psi^*(x)}(da), \qquad \forall x \in \zeta^c. \tag{5.16}$$

Definition 5.4. *Let us introduce the occupation measure $\mathbf{M}_\gamma^\sigma$ of a policy σ for the undiscoutned DTMDP model $\{S_\infty, A, p, \gamma\}$ as a measure on $\mathcal{B}(S \times A)$ defined by*

$$\mathbf{M}_\gamma^\sigma(\Gamma_S \times \Gamma_A) := \sum_{n=0}^\infty \mathbf{P}_\gamma^\sigma(X_n \in \Gamma_S, A_{n+1} \in \Gamma_A) \tag{5.17}$$

for each $\Gamma_S \in \mathcal{B}(S)$ *and* $\Gamma_A \in \mathcal{B}(A)$. *Here, as before, the transition probability* p *is defined by* (3.4) *and* (3.5).

The next statement is a consequence of Theorem 4.1 and its proof.

Corollary 5.2. *Suppose Condition 4.1 is satisfied. Consider a feasible policy* $\pi = (\pi_n) \in \Pi$ *with a finite value for the CTMDP problem* (2.2) *such that*

$$\pi_{n+1}(da|x_0, \theta_1, \dots, x_n, s) = \delta_{\psi^*(x_n)}(da) \tag{5.18}$$

whenever $x_n \in \zeta^c$. *Then there is a Markov policy* σ *for the DTMDP* $\{S_\infty, A, p, \gamma\}$ *such that*

$$\sigma_{n+1}(da|x_n) = \delta_{\psi^*(x_n)}(da) \tag{5.19}$$

for each $x_n \in \zeta^c$, *and*

$$q_x(a)\eta^\pi_\gamma(dx \times da) = \mathbf{M}^\sigma_\gamma(dx \times da) \tag{5.20}$$

on $\mathcal{B}(\zeta \times A)$. *Here the mapping* ψ^* *is the fixed one satisfying* (5.12) *and* (5.13).

Proof. Inspecting the proof of Theorem 4.1, one can see that any Markov policy $\sigma = (\sigma_n)$ for the DTMDP model $\{S_\infty, A, p\}$ with $\sigma_{n+1}(da|x)$ satisfying (4.14) and (4.17) for each $n = 0, 1, \dots$ fulfils the conditions of the statement of Theorem 4.1; and there exists at least one such policy, which we consider now. On $\mathcal{B}(\zeta^c \setminus S_2 \times A)$, (4.17) reads that for each $n = 0, 1, \dots$

$$\begin{aligned} q_x(a)m^\pi_{\gamma,n}(dx \times da) &= \left(\int_A m^\pi_{\gamma,n}(dx \times db)q_x(b)\right)\sigma_{n+1}(da|x) \\ \Leftrightarrow \quad q_x(a)m^\pi_{\gamma,n}(dx \times A)\delta_{\psi^*(x)}(da) &= m^\pi_{\gamma,n}(dx \times A)q_x\big(\psi^*(x)\big)\sigma_{n+1}(da|x), \end{aligned}$$

where the equivalence is by (5.18). Therefore, one can always put $\sigma_{n+1}(da|x) = \delta_{\psi^*(x)}(da)$ for each $x \in \zeta^c \setminus S_2$ without violating (4.17). This together with (4.14) shows that the policy σ satisfies (5.19); recall (5.13). From the discussion in the beginning of this proof, this policy σ satisfies (4.15), by summing up both sides of which with respect to n, we see that (5.20) is also fulfilled. The corollary is now proved. □

We end this section with the next lemma.

Lemma 5.5. *Suppose Condition 4.1 is satisfied. Let some* σ *be a policy for the DTMDP model* $\{S_\infty, A, p, \gamma\}$ *such that*

$$\int_{S\times A} \mathbf{M}^\sigma_\gamma(dx \times da)\frac{c_i(x,a)}{q_x(a)} < \infty$$

for each $i = 0, 1, \dots, N$. *Suppose that there exists a stationary policy* σ^S *for the DTMDP model* $\{S_\infty, A, p, \gamma\}$ *satisfying* $\mathbf{M}^\sigma_\gamma(dx \times da) = \mathbf{M}^\sigma_\gamma(dx \times A)\sigma^S(da|x)$ *on* $\mathcal{B}(\zeta \times A)$, *and* $\sigma^S(da|x) =$

$\delta_{\psi^*(x)}(da)$ *for each* $x \in \zeta^c$. *Then*

$$\mathbf{M}_\gamma^{\sigma^S}(dx \times da) \le \mathbf{M}_\gamma^{\sigma}(dx \times da)$$

on $\mathcal{B}(\zeta \times A)$.

Proof. According to Proposition 4.1; see especially (4.7), and Proposition 9.10 of [5],

$$\inf_{\pi \in \Pi_{\mathrm{DM}}} E_x^\pi \left[\int_0^\infty \int_A \sum_{i=0}^N c_i(\xi_t, a) \pi(da|\omega, t)\, dt \right] = \inf_\sigma \mathbf{E}_x^\sigma \left[\sum_{n=0}^\infty \sum_{i=0}^N \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})} \right]$$

for each $x \in S$. Thus,

$$\zeta = \left\{ x \in S : \inf_\sigma \mathbf{E}_x^\sigma \left[\sum_{n=0}^\infty \sum_{i=0}^N \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})} \right] > 0 \right\}.$$

Now one can apply Theorem 3.3 of [10] for the statement. We remark that in [10], only nonnegative finitely valued cost functions were considered for the concerned DTMDP model. A careful inspection of the reasonings therein reveal that all the cited statements from [10] in this paper survive when the cost functions are nonnegative extended real-valued. □

6. Optimality result

The main objective of this section is to show the existence of a stationary optimal policy for the CTMDP problem (2.2); see Theorem 6.2 below. In the process, we also justify the reduction of the CTMDP problem (2.2) to the DTMDP problem (3.3) under Conditions 4.1 and 6.1; see Remark 6.1 below. To this end, we firstly show the sufficiency of stationary policies for the CTMDP problem (2.2); see Theorem 6.1.

Theorem 6.1. *Suppose Condition 4.1 is satisfied, and consider a feasible policy* π *with a finite value for the CTMDP problem* (2.2) *such that for each* $n = 0, 1, \ldots,$

$$\pi_{n+1}(da|x_0, \theta_1, \ldots, x_n, s) = \delta_{\psi^*(x_n)}(da) \tag{6.1}$$

whenever $x_n \in \zeta^c$. *Then the stationary policy* φ_π *for the CTMDP problem* (2.2) *coming from Corollary 5.1 satisfies*

$$E_\gamma^{\varphi_\pi} \left[\int_0^\infty \int_A c_i(\xi_t, a) \varphi_\pi(da|\xi_t)\, dt \right] \le E_\gamma^\pi \left[\int_0^\infty \int_A c_i(\xi_t, a) \pi(da|\omega, t)\, dt \right] \tag{6.2}$$

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

Proof. The proof goes in several steps.

Step 1. We show that the stationary policy φ_π satisfies that

$$\varphi_\pi\big(B(x)|x\big) < 1$$

for almost all $x \in S_1$ with respect to $\eta^\pi_\gamma(dx \times A)$, where $B(x)$ is given by (4.3), and S_1 is given by (4.2).

It suffices to prove the above claim for the case of

$$\eta^\pi_\gamma(S_1 \times A) > 0 \tag{6.3}$$

as follows.

Note that

$$\begin{aligned}
\infty > \int_{S_1\times A} \eta^\pi_\gamma(dx\times da)\sum_{i=0}^N c_i(x,a) &= \int_{S_1}\eta^\pi_\gamma(dx\times A)\int_A \varphi_\pi(da|x)\sum_{i=0}^N c_i(x,a)\\
&= \sum_{n=0}^\infty E^\pi_\gamma\left[E^\pi_\gamma\left[\int_{t_n}^{t_{n+1}}\int_{S_1} I\{x_n\in dx\}\int_A \varphi_\pi(da|x)\sum_{i=0}^N c_i(x,a)\,dt|x_0,\theta_1,\dots,x_n\right]\right] \qquad (6.4)\\
&= \sum_{n=0}^\infty E^\pi_\gamma\left[\int_{S_1} I\{x_n\in dx\}\int_A \varphi_\pi(da|x)\sum_{i=0}^N c_i(x,a)E^\pi_\gamma[\theta_{n+1}|x_0,\theta_1,\dots,x_n]\right].
\end{aligned}$$

Suppose for contradiction that

$$\varphi_\pi\big(B(x)|x\big) = 1 \tag{6.5}$$

on a measurable subset $\Gamma_1 \subseteq S_1$ of positive measure with respect to $\eta^\pi_\gamma(dx \times A)$. It holds that

$$\int_A \varphi_\pi(da|x)\sum_{i=0}^N c_i(x,a) \geq \int_{B(x)} \varphi_\pi(da|x)\sum_{i=0}^N c_i(x,a) > 0 \tag{6.6}$$

for each $x \in \Gamma_1 \subseteq S_1$, where the last inequality is by (4.3) and (4.2).

According to (6.3), there exists some $n = 0, 1, \dots$ such that

$$P^\pi_\gamma(x_n \in \Gamma_1) > 0;$$

and for this n, it must hold that

$$E^\pi_\gamma[\theta_{n+1}|x_0,\theta_1,\dots,x_n] < \infty \tag{6.7}$$

for almost all $\omega \in \{\omega\in\Omega : x_n(\omega)\in\Gamma_1\}$ with respect to $P^\pi_\gamma(d\omega)$, for otherwise this together with (6.6) would contradict the first inequality of (6.4).

The definition of $B(x)$ given by (4.3) and the inequality (6.7) imply that

$$\eta^\pi_\gamma\big(\{(x,a) : x\in\Gamma_1, a\in A\setminus B(x)\}\big) > 0, \tag{6.8}$$

where the set in the bracket is measurable because so is the set $\{(x,a) : x \in \Gamma_1, a \in B(x)\}$ according to e.g., Theorem 3.1 of Feinberg *et al.* [15]. Since

$$\begin{aligned}
&\int_{\Gamma_1} \varphi_\pi\big(A \setminus B(x)|x\big)\eta_\gamma^\pi(dx \times A) \\
&\quad = \int_{\Gamma_1 \cap \zeta} \varphi_\pi\big(A \setminus B(x)|x\big)\eta_\gamma^\pi(dx \times A) + \int_{\Gamma_1 \cap (\zeta^c)} \varphi_\pi\big(A \setminus B(x)|x\big)\eta_\gamma^\pi(dx \times A) \\
&\quad = \eta_\gamma^\pi\big(\big\{(x,a) : x \in \Gamma_1, a \in A \setminus B(x)\big\}\big),
\end{aligned}$$

which holds by (5.15) and (6.1), the relation (6.8) implies that $\varphi_\pi(A \setminus B(x)|x) > 0$ on some measurable subset of $\Gamma_2 \subseteq \Gamma_1$ of positive measure with respect to $\eta_\gamma^\pi(dx \times A)$. This is a desired contradiction against the relation in (6.5). Step 1 is completed.

Step 2. Consider the policy σ for the DTMDP model $\{S_\infty, A, p, \gamma\}$ from Corollary 5.2, and define the stationary policy σ^S for the DTMDP model $\{S_\infty, A, p, \gamma\}$ by

$$\sigma^S(da|x) = \delta_{\psi^*(x)}(da)$$

for all $x \in \zeta^c$ and for all $x \in \zeta$ satisfying $\int_A q_x(a)\varphi_\pi(da|x) = 0$; and

$$\sigma^S(da|x) := \frac{q_x(a)\varphi_\pi(da|x)}{\int_A q_x(a)\varphi_\pi(da|x)} \tag{6.9}$$

for all $x \in \zeta$ such that $\int_A q_x(a)\varphi_\pi(da|x) > 0$. Recall that ψ^* is the fixed measurable mapping satisfying (5.12) and (5.13). We verify that

$$\mathbf{M}_\gamma^\sigma(dx \times A)\sigma^S(da|x) = \mathbf{M}_\gamma^\sigma(dx \times da)$$

on $\mathcal{B}(\zeta \times A)$; recall (5.17) for the definition of $\mathbf{M}_\gamma^\sigma$. Throughout the proof of this theorem, the policies σ and σ^S are understood as here.

Indeed, on $\mathcal{B}(\zeta \times A)$, it holds that

$$\begin{aligned}
\mathbf{M}_\gamma^\sigma(dx \times A)\sigma^S(da|x) &= \left(\int_A \eta_\gamma^\pi(dx \times db)q_x(b)\right)\sigma^S(da|x) \\
&= \left(\int_A \eta_\gamma^\pi(dx \times A)\varphi_\pi(db|x)q_x(b)\right)\sigma^S(da|x) \\
&= \eta_\gamma^\pi(dx \times A)\left(\int_A \varphi_\pi(db|x)q_x(b)\right)\frac{q_x(a)\varphi_\pi(da|x)}{\int_A q_x(a)\varphi_\pi(da|x)} \\
&= \eta_\gamma^\pi(dx \times A)q_x(a)\varphi_\pi(da|x) \\
&= \mathbf{M}_\gamma^\sigma(dx \times da),
\end{aligned}$$

where the first and the last equalities are by (5.20), the second equality is by (5.15), the third and forth equalities are by (6.9); and the fact that $\int_A q_x(a)\varphi_\pi(da|x) > 0$ for almost all $x \in \zeta$, which

in turn follows from the facts that $\int_A q_x(a)\psi_\pi(da|x) > 0$ for almost all $x \in S_1$ with respect to $\eta^\pi_\gamma(dx \times A)$ as established in Step 1; $\int_A q_x(a)\psi_\pi(da|x) > 0$ for all $x \in S_3$ by (4.2); and the relation $\zeta \subseteq S_1 \cup S_3$. Step 2 is thus completed.

Step 3. We verify that

$$\mathbf{M}^{\sigma^S}_\gamma(dx \times A)\sigma^S(da|x) = \mathbf{M}^{\sigma^S}_\gamma(dx \times da) \le \mathbf{M}^{\sigma}_\gamma(dx \times da) \tag{6.10}$$

on $\mathcal{B}(\zeta \times A)$ as follows.

The equality in (6.10) holds because the policy σ^S is stationary and (5.17). For the inequality in (6.10), we observe that

$$\begin{aligned}
&\int_{S\times A} \mathbf{M}^{\sigma}_\gamma(dx \times da)\frac{c_i(x,a)}{q_x(a)} \\
&= \int_{\zeta\times A} \mathbf{M}^{\sigma}_\gamma(dx \times da)\frac{c_i(x,a)}{q_x(a)} + \int_{\zeta^c} \mathbf{M}^{\sigma}_\gamma(dx \times A)\frac{c_i(x,\psi^*(x))}{q_x(\psi^*(x))} \\
&= \int_{\zeta\times A} \eta^\pi_\gamma(dx \times da)q_x(a)\frac{c_i(x,a)}{q_x(a)} \le \int_{\zeta\times A} \eta^\pi_\gamma(dx \times da)c_i(x,a) \\
&\le \int_{S\times A} c_i(x,a)\eta^\pi_\gamma(dx \times da) < \infty
\end{aligned}$$

for each $i = 0, 1, \ldots, N$, where the first equality is by (5.19), the second equality is by (5.14), the first inequality is by that $q_x(a)\frac{c_i(x,a)}{q_x(a)} \le c_i(x,a)$; recall the convention of $\frac{0}{0} = 0$ and $0 \cdot \infty = 0$, and the last inequality is by that the policy π is feasible with a finite value for problem (2.2). With this inequality and the equality of (6.10) in hand, we see that the conditions of Lemma 5.5 are satisfied, following from which, the inequality of (6.10) holds. Step 3 is completed.

Step 4. Let us introduce the set

$$\zeta_\pi := \left\{x \in \zeta : \int_A q_x(a)\varphi_\pi(da|x) = 0\right\}, \tag{6.11}$$

which is measurable. We establish

$$\eta^{\varphi_\pi}_\gamma(dx \times da)q_x(a) = \mathbf{M}^{\sigma^S}_\gamma(dx \times da) \tag{6.12}$$

on $\mathcal{B}((\zeta \setminus \zeta_\pi) \times A)$.

To this end, we show by induction the more detailed relation

$$M^{n,\varphi_\pi}_\gamma(dx \times da) = \mathbf{P}^{\sigma^S}_\gamma(X_n \in dx, A_{n+1} \in da) \tag{6.13}$$

on $\mathcal{B}(\zeta \setminus \zeta_\pi \times A)$ for each $n = 0, 1, \ldots$ as follows.

Consider $n = 0$. Then on $\mathcal{B}(\zeta \setminus \zeta_\pi)$,

$$M^{0,\varphi_\pi}_\gamma(dx \times A) = \gamma(dx) = \mathbf{P}^{\sigma^S}_\gamma(X_0 \in dx), \tag{6.14}$$

where the first equality is by (4.18) and the fact that $\zeta \subseteq S \setminus S_2$. Now on $\mathcal{B}(\zeta \setminus \zeta_\pi \times A)$,

$$\begin{aligned}
&\mathbf{P}_\gamma^{\sigma^S}(X_0 \in dx, A_1 \in da) \\
&= \mathbf{P}_\gamma^{\sigma^S}(X_0 \in dx)\sigma^S(da|x) = M_\gamma^{0,\varphi_\pi}(dx \times A)\sigma^S(da|x) \\
&= \int_A E_\gamma^{\varphi_\pi}\left[\int_0^{t_1} q_x(b)\varphi_\pi(db|x) I\{x_0 \in dx\}\, dt\right] \frac{q_x(a)\varphi_\pi(da|x)}{\int_A q_x(a)\varphi_\pi(da|x)} \qquad (6.15) \\
&= E_\gamma^{\varphi_\pi}\left[\int_0^{t_1} q_x(a)\varphi_\pi(da|x) I\{x_0 \in dx\}\, dt\right] \\
&= M_\gamma^{0,\varphi_\pi}(dx \times da),
\end{aligned}$$

where the second equality is by (6.14), the third equality is by (6.9); remember that

$$\int_A q_x(a)\varphi_\pi(da|x) > 0, \qquad \forall x \in \zeta \setminus \zeta_\pi.$$

Assume (6.13) holds on $\mathcal{B}(\zeta \setminus \zeta_\pi \times A)$ for all $n \le k$, and consider the case of $n = k+1$. On the one hand, on $\mathcal{B}(\zeta \setminus \zeta_\pi)$ it holds that

$$\begin{aligned}
&\mathbf{P}_\gamma^{\sigma^S}(X_{k+1} \in dx) \\
&= \int_{S\times A} \frac{\tilde{q}(dx|y,a)}{q_y(a)} \mathbf{P}_\gamma^{\sigma^S}(X_k \in dy, A_{k+1} \in da) \\
&= \int_{S\times A} \frac{\tilde{q}(dx|y,a)}{q_y(a)} M_\gamma^{k,\varphi_\pi}(dy \times da) = \int_{S\times A} \frac{\tilde{q}(dx|y,a)}{q_y(a)} q_y(a) m_{\gamma,k}^{\varphi_\pi}(dy \times da) \\
&= \int_{S\times A} \tilde{q}(dx|y,a) m_{\gamma,k}^{\varphi_\pi}(dy \times da) \\
&= E_\gamma^{\varphi_\pi}\left[\int_A \tilde{q}(dx|x_k,a)\varphi_\pi(da|x_k) E_\gamma^{\varphi_\pi}[\theta_{k+1}|x_0,\theta_1,\ldots,x_k]\right] \\
&= E_\gamma^{\varphi}\left[\frac{\int_A \tilde{q}(dx|x_k,a)\varphi_\pi(da|x_k)}{\int_A q_{x_k}(a)\varphi_\pi(da|x_k)}\right],
\end{aligned}$$

where the second equality is by the inductive supposition, the forth equality is by that $\frac{\tilde{q}(dx|y,a)}{q_y(a)} q_y(a) = \tilde{q}(dx|y,a)$ no matter whether $q_y(a)$ vanishes or not, and the last equality holds due to the convention of $\frac{0}{0} = 0$. On the other hand, on $\mathcal{B}(\zeta \setminus \zeta_\pi)$,

$$\begin{aligned}
&M_{k+1,\gamma}^{\varphi_\pi}(dx \times A) \\
&= E_\gamma^{\varphi_\pi}\left[E_\gamma^{\varphi_\pi}\left[I\{x_{k+1} \in dx\}\right.\right.
\end{aligned}$$

$$\times E_\gamma^{\varphi_\pi}\left[\int_0^{\theta_{k+2}}\int_A q_{x_{k+1}}(a)\varphi_\pi(da|x_{k+1})|x_0,\dots,\theta_{k+1},x_{k+1}\right]\Big|\,x_0,\dots,x_k,\theta_{k+1}\Big]\Big]$$
$$= E_\gamma^{\varphi_\pi}\left[E_\gamma^{\varphi_\pi}\left[I\{x_{k+1}\in dx\}|x_0,\theta_1,\dots,x_k,\theta_{k+1}\right]\right]$$
$$= E_\gamma^{\varphi_\pi}\left[\frac{\int_A \tilde{q}(dx|x_k,a)\varphi_\pi(da|x_k)}{\int_A(q_{x_k}(a))\varphi_\pi(da|x_k)}\right].$$

Thus,

$$M_{k+1,\gamma}^{\varphi_\pi}(dx\times A)=\mathbf{P}_\gamma^{\sigma^S}(X_{k+1}\in dx)$$

on $\mathcal{B}(\zeta\setminus\zeta_\pi)$. Based on this, a similar calculation as the one for (6.15) leads to

$$M_{k+1,\gamma}^{\varphi_\pi}(dx\times da)=\mathbf{P}_\gamma^{\sigma^S}(X_{k+1}\in dx, A_{k+1}\in da)$$

on $\mathcal{B}(\zeta\setminus\zeta_\pi)\times A)$. Hence, (6.13) is shown by induction, and (6.12) follows. Step 4 is completed.

Step 5. We show that

$$\eta_\gamma^{\varphi_\pi}(dx\times da)\le\eta_\gamma^{\pi}(dx\times da) \tag{6.16}$$

on $\mathcal{B}(\zeta\setminus\zeta_\pi\times A)$.

Indeed, by (6.10) and (6.12) as established in Steps 3 and 4, we see

$$\eta_\gamma^{\varphi_\pi}(dx\times da)q_x(a)\le\mathbf{M}_\gamma^{\sigma}(dx\times da)$$

on $\mathcal{B}(\zeta\setminus\zeta_\pi\times A)$, which together with (5.20) further leads to

$$\eta_\gamma^{\varphi_\pi}(dx\times da)q_x(a)\le\eta_\gamma^{\pi}(dx\times da)q_x(a) \tag{6.17}$$

on $\mathcal{B}(\zeta\setminus\zeta_\pi\times A)$. Now on $\mathcal{B}(\zeta\setminus\zeta_\pi)$,

$$E_\gamma^{\varphi_\pi}\left[\int_0^\infty\int_A q_x(a)\varphi_\pi(da|x)I\{\xi_t\in dx\}\,dt\right]$$
$$=\left(\int_A\varphi_\pi(da|x)q_x(a)\right)\eta_\gamma^{\varphi_\pi}(dx\times A)=\int_A\eta_\gamma^{\varphi_\pi}(dx\times da)q_x(a)\le\int_A\eta_\gamma^{\pi}(dx\times da)q_x(a)$$
$$=\left(\int_A q_x(a)\varphi_\pi(da|x)\right)\eta_\gamma^{\pi}(dx\times A),$$

where the inequality is by (6.17), and the last equality is by (5.15). Since

$$\int_A q_x(a)\varphi_\pi(da|x)>0$$

for all $x\in\zeta\setminus\zeta_\pi$, we infer from the above inequality for that

$$\eta_\gamma^{\varphi_\pi}(dx\times A)\le\eta_\gamma^{\pi}(dx\times A)$$

on $\mathcal{B}(\zeta \setminus \zeta_\pi)$, from which (6.16) holds on $\mathcal{B}(\zeta \setminus \zeta_\pi \times A)$; recall (5.15). Step 5 is completed.

Step 6. We show that

$$\eta_\gamma^{\varphi_\pi}(\zeta_\pi \times A) = 0. \tag{6.18}$$

Suppose for contradiction that

$$\eta_\gamma^{\varphi_\pi}(\zeta_\pi \times A) > 0. \tag{6.19}$$

Note that $\zeta_\pi \subseteq S_1$, where ζ_π is given by (6.11); recall that $\zeta \subseteq S_1 \cup S_3$ and the definition of S_3. Therefore, the statement established in Step 1 implies that

$$\eta_\gamma^{\pi}(\zeta_\pi \times A) = 0. \tag{6.20}$$

Therefore, $\gamma(\zeta_\pi) = 0$. Now following from (6.19), there exists some $\Gamma \in \mathcal{B}(S \setminus \zeta_\pi)$ satisfying that

$$\int_A \tilde{q}(\zeta_\pi | x, a)\varphi_\pi(da|x) > 0 \tag{6.21}$$

for all $x \in \Gamma$, and

$$\eta_\gamma^{\varphi_\pi}(\Gamma \times A) > 0. \tag{6.22}$$

Note that according to (5.16), the definition of the set ζ given by (5.10), and (5.12), we see that $\tilde{q}(\zeta | x, \psi^*(x)) = 0$ for each $x \in \zeta^c$. Since $\zeta_\pi \subseteq \zeta$, we see $\tilde{q}(\zeta_\pi | x, \psi^*(x)) = 0$ for each $x \in \zeta^c$. Consequently, we have

$$\Gamma \in \mathcal{B}(\zeta \setminus \zeta_\pi)$$

for otherwise it would contradict (6.21). This fact, (6.22) and (6.16) as established in Step 5 show that

$$\eta_\gamma^{\pi}(\Gamma \times A) > 0; \qquad \eta_\gamma^{\pi}(dx \times A) \geq \eta_\gamma^{\varphi_\pi}(dx \times A) \qquad \text{on } \mathcal{B}(\Gamma). \tag{6.23}$$

Now

$$\int_{\Gamma \times A} \eta_\gamma^{\pi}(dx \times da)\tilde{q}(\zeta_\pi | x, a) = \int_\Gamma \eta_\gamma^{\pi}(dx \times A) \int_A \tilde{q}(\zeta_\pi | x, a)\varphi_\pi(da|x) > 0,$$

where the first equality is by (5.15), and the last inequality is by (6.23). Thus,

$$E_\gamma^{\pi}\left[\int_0^\infty \int_A \tilde{q}(\zeta_\pi | \xi_t, a)\pi(da|\omega, t) I\{\xi_t \in \Gamma\}\, dt\right] > 0.$$

It follows from this inequality and the construction of the CTMDP that $\eta_\gamma^{\pi}(\zeta_\pi \times A) > 0$, which is a contradiction against (6.20). Hence, (6.18) holds. Step 6 is completed.

Step 7. We prove the statement of the theorem now. It holds that for each $i = 0, 1, \ldots, N$,

$$\begin{aligned}
&\int_{S\times A} \eta_\gamma^\pi(dx \times da)c_i(x,a) \\
&= \int_{\zeta\setminus\zeta_\pi\times A} \eta_\gamma^\pi(dx \times da)c_i(x,a) + \int_{\zeta^c\times A} \eta_\gamma^\pi(dx \times da)c_i\big(x, \psi^*(x)\big) \\
&\quad + \int_{\zeta_\pi\times A} \eta_\gamma^\pi(dx \times da)c_i(x,a) \\
&\geq \int_{\zeta\setminus\zeta_\pi\times A} \eta_\gamma^{\varphi_\pi}(dx \times da)c_i(x,a) + \int_{\zeta^c\times A} \eta_\gamma^{\varphi_\pi}(dx \times da)c_i\big(x, \psi^*(x)\big) \\
&\quad + \int_{\zeta_\pi\times A} \eta_\gamma^{\varphi_\pi}(dx \times da)c_i(x,a) \\
&= \int_{S\times A} \eta_\gamma^{\varphi_\pi}(dx \times da)c_i(x,a),
\end{aligned}$$

where the first equality is by (6.1), and the inequality is by (5.14), (5.16), (6.16), and (6.18). Thus, (6.2) is proved. □

Corollary 6.1. *Suppose Condition 4.1 is satisfied, and consider a feasible policy π with a finite value for the CTMDP problem (2.2) satisfying (6.1) as in the statement of Theorem 6.1. Then there exists a stationary policy ϕ_π such that*

$$\phi_\pi\big(B(x)|x\big) = 0 \tag{6.24}$$

for each $x \in S_1 \setminus \hat{S}_1$ provided that $S_1 \setminus \hat{S}_1 \neq \varnothing$,

$$\phi_\pi(da|x) = \delta_{\psi^*(x)}(da) \tag{6.25}$$

for each $x \in \zeta^c$ whenever $\zeta^c \neq \varnothing$, and

$$E_\gamma^{\phi_\pi}\left[\int_0^\infty \int_A c_i(\xi_t, a)\phi_\pi(da|\xi_t)\,dt\right] \leq E_\gamma^{\pi}\left[\int_0^\infty \int_A c_i(\xi_t, a)\pi(da|\omega, t)\,dt\right]$$

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

Proof. Let the stationary policy φ_π be as in the statement of Theorem 6.1. Assume that $S_1 \setminus \hat{S}_1 \neq \varnothing$; the other case is simpler. For each $x \in S_1 \setminus \hat{S}_1$, $A \setminus B(x) \neq \varnothing$; this is by the definitions of $B(x)$, S_1 and $\hat{S}_1$; see (4.3) and (4.2). By Proposition 7.33 of [5], there is a measurable mapping $\hat{\psi}$ from $S_1 \setminus \hat{S}_1$ to A such that

$$\sup_{a\in A} q_x(a) = q_x\big(\hat{\psi}(x)\big) > 0$$

for each $x \in S_1 \setminus \hat{S}_1$, where the inequality follows from the fact that $\sup_{a\in A} q_x(a) = \max_{a\in A} q_x(a) = \max_{a\in A\setminus B(x)} q_x(a) > 0$; recall the definition of $B(x)$ as given by (4.3). Observe that

$$\{x \in S_1 \setminus \hat{S}_1 : \varphi_\pi(A \setminus B(x)|x) = 0\} = \{x \in (S_1 \setminus \hat{S}_1) \cap \zeta : \varphi_\pi(A \setminus B(x)|x) = 0\} \quad (6.26)$$

by (5.12) and the definition of S_1. Now if

$$\{x \in S_1 \setminus \hat{S}_1 : \varphi_\pi(A \setminus B(x)|x) = 0\} \neq \varnothing,$$

then we modify the definition of φ_π by putting (with slight abuse of notations by using φ_π for both the original and the modified policies) $\varphi_\pi(da|x) := \delta_{\hat{\psi}(x)}(da)$ for each $x \in \{x \in (S_1 \setminus \hat{S}_1) \cap \zeta : \varphi_\pi(A \setminus B(x)|x, a) = 0\}$. Since

$$\eta_\gamma^\pi(\{x \in S_1 \setminus \hat{S}_1 : \varphi_\pi(A \setminus B(x)|x) = 0\}) = 0$$

as established in Step 1 of the proof of Theorem 6.1, the resulting stationary policy φ_π still satisfies (5.15) and (5.16); recall (6.26). Therefore, Theorem 6.1 remains applicable to this modified policy. For this reason, in the rest of this proof, we suppose without loss of generality that

$$\{x \in S_1 \setminus \hat{S}_1 : \varphi_\pi(A \setminus B(x)|x) = 0\} = \varnothing. \quad (6.27)$$

Now define a stationary policy ϕ_π by

$$\phi_\pi(da|x) := \frac{\varphi_\pi(da \cap (A \setminus B(x))|x)}{\varphi_\pi((A \setminus B(x))|x)}$$

for each $x \in S_1 \setminus \hat{S}_1$, and

$$\phi_\pi(da|x) := \varphi_\pi(da|x)$$

elsewhere. Observe that ϕ_π defined in the above is indeed a stochastic kernel; this follows from the fact that $\{(x, a) : q_x(a) = 0\} = \{(x, a) : a \in B(x)\}$ is measurable, which is by Theorem 3.1 of [15]; see also Corollary 18.8 of [1], and Proposition 7.29 of [5]. The relation (6.25) holds for this policy ϕ_π because of its definition and (5.16); observe that for each $x \in (S_1 \setminus \hat{S}_1) \cap (\zeta^c)$, it holds that $\psi^*(x) \notin B(x)$.

Direct calculations show that for each $x \in S$,

$$\frac{\int_A \tilde{q}(dy|x, a)\phi_\pi(da|x)}{\int_A q_x(a)\phi_\pi(da|x)} = \frac{\int_A \tilde{q}(dy|x, a)\varphi_\pi(da|x)}{\int_A q_x(a)\varphi_\pi(da|x)}.$$

Also observe that for each $x \in S_1 \setminus \hat{S}_1$ and $i = 0, 1, \ldots, N$,

$$\begin{aligned}
&\int_0^\infty \int_A c_i(x, a)\phi_\pi(da|x)e^{-\int_A q_x(a)\phi_\pi(da|x)t}\,dt \\
&\quad \le \int_0^\infty \frac{\int_A c_i(x, a)\varphi_\pi(da|x)}{\varphi_\pi(A \setminus B(x)|x)} e^{-\int_A q_x(a)\varphi_\pi(da|x)t \frac{1}{\varphi_\pi(A\setminus B(x)|x)}}\,dt
\end{aligned}$$

$$= \int_A c_i(x,a)\varphi_\pi(da|x)\frac{1}{\int_A q_x(a)\varphi_\pi(da|x)}$$
$$= \int_0^\infty \int_A c_i(x,a)\varphi_\pi(da|x)e^{-\int_A q_x(a)\varphi_\pi(da|x)t}\,dt;$$

remember, $\int_A q_x(a)\varphi_\pi(da|x) > 0$ for each $x \in S_1 \setminus \hat{S}_1$ by (6.27). In other words, under the stationary policy ϕ_π, given the current state $x \in S$, the (conditional) distribution of the next jump-in state is the same as the one under the stationary policy φ_π, and the total (conditional) expected cost during the current sojourn time is not larger than the one under φ_π. Since both policies φ_π and ϕ_π are stationary, this and Theorem 6.1 prove the statement. □

Corollary 6.2. *Suppose Condition 4.1 is satisfied, and consider a feasible policy π with a finite value for the CTMDP problem (2.2) satisfying (6.1) as in the statement of Theorem 6.1. Then there exists a stationary policy σ_π^S for the DTMDP model $\{S_\infty, A, p, \gamma\}$ such that for each $i = 0, 1, \ldots, N$,*

$$\mathbf{E}_\gamma^{\sigma_\pi^S}\left[\sum_{n=0}^\infty \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] \le E_\gamma^\pi\left[\int_0^\infty \int_A c_i(\xi_t, a)\pi(da|\omega,t)\,dt\right].$$

Proof. Let ϕ_π be the stationary policy for the CTMDP model coming from Corollary 6.1. By Theorem 4.1, there is a Markov policy say $\sigma_\pi^M = (\sigma_\pi^M{}_n)$ for the DTMDP model $\{S_\infty, A, p, \gamma\}$ satisfying, for each $n = 0, 1, \ldots,$

$$M_\gamma^{n,\phi_\pi}(dx \times da) = \mathbf{P}_\gamma^{\sigma_\pi^M}(X_n \in dx, A_{n+1} \in da) \tag{6.28}$$

on $\mathcal{B}(S \setminus S_2 \times A)$, and

$$\sigma_\pi^M{}_{n+1}(da|x) = \delta_{f^*(x)}(da) \tag{6.29}$$

for each $x \in S_2$ whenever $S_2 \neq \varnothing$.

Now for each $i = 0, 1, \ldots, N$, it holds that

$$\begin{aligned} &E_\gamma^{\phi_\pi}\left[\int_0^\infty \int_A c_i(\xi_t, a)\phi_\pi(da|\xi_t)\,dt\right] \\ &= \sum_{n=0}^\infty \int_{S\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx \times da) \\ &= \sum_{n=0}^\infty \left\{\int_{S_1\setminus\hat{S}_1\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx \times da) + \int_{\hat{S}_1\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx \times da)\right. \\ &\quad \left. + \int_{S_2\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx \times da) + \int_{S_3\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx \times da)\right\}. \end{aligned} \tag{6.30}$$

The first term in the summand in the last line of the above equality can be written as follows:

$$\begin{aligned}\int_{S_1\setminus\hat{S}_1\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx\times da) &= \int_{S_1\setminus\hat{S}_1}\int_A c_i(x,a)\phi_\pi(da|x)m_{\gamma,n}^{\phi_\pi}(dx\times A)\\ &= \int_{S_1\setminus\hat{S}_1}\int_A \frac{c_i(x,a)}{q_x(a)}q_x(a)\phi_\pi(da|x)m_{\gamma,n}^{\phi_\pi}(dx\times A)\\ &= \int_{S_1\setminus\hat{S}_1\times A}\frac{c_i(x,a)}{q_x(a)}M_\gamma^{n,\phi_\pi}(dx\times da)\\ &= \int_{S_1\setminus\hat{S}_1\times A}\frac{c_i(x,a)}{q_x(a)}\mathbf{P}_\gamma^{\sigma_\pi^M}(X_n\in dx, A_{n+1}\in da),\end{aligned}$$

where the second equality holds because of (6.24), and the third equality is by the definitions of M_γ^{n,ϕ_π} and $m_{\gamma,n}^{\phi_\pi}$, and the last equality is by (6.28). For the second term in the summand in the last line of (6.30), we have

$$\int_{\hat{S}_1\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx\times da) = \int_{\hat{S}_1\times A}\frac{c_i(x,a)}{q_x(a)}\mathbf{P}_\gamma^{\sigma_\pi^M}(X_n\in dx, A_{n+1}\in da),$$

where the equality holds because

$$m_{\gamma,n}^{\phi_\pi}(\hat{S}_1\times A) = 0 = \mathbf{P}_\gamma^{\sigma_\pi^M}(X_n\in\hat{S}_1)$$

with the first equality being by Lemma 4.4 (see (4.11) therein) applied to ϕ_π, which is feasible with a finite value for problem (2.2) for it outperforms the policy π by Corollary 6.1, and the second equality being valid by (6.28) and that $M_\gamma^{n,\phi_\pi}(dx\times da) = q_x(a)m_{\gamma,n}^{\phi_\pi}(dx\times da)$. For the third term in the summand in the last line of (6.30),

$$\begin{aligned}\int_{S_2\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx\times da) &= \int_{S_2} c_i\big(x,\psi^*(x)\big)m_{\gamma,n}^{\phi_\pi}(dx\times A) = 0\\ &= \int_{S_2}\frac{c_i(x,\psi^*(x))}{q_x(\psi^*(x))}\mathbf{P}_\gamma^{\sigma_\pi^M}(X_n\in dx)\\ &= \int_{S_2\times A}\frac{c_i(x,a)}{q_x(a)}\mathbf{P}_\gamma^{\sigma_\pi^M}(X_n\in dx, A_{n+1}\in da),\end{aligned}$$

where the first equality is by (6.25); recall that $S_2\subseteq\zeta^c$, the second and third equalities are by (5.14), and the last equality is by (6.29) and (5.13). Finally, for the last term in the summand of (6.30), it holds that

$$\begin{aligned}\int_{S_3\times A} c_i(x,a)m_{\gamma,n}^{\phi_\pi}(dx\times da) &= \int_{S_3\times A}\frac{c_i(x,a)}{q_x(a)}q_x(a)m_{\gamma,n}^{\phi_\pi}(dx\times da)\\ &= \int_{S_3\times A}\frac{c_i(x,a)}{q_x(a)}M_\gamma^{n,\phi_\pi}(dx\times da)\end{aligned}$$

$$= \int_{S_3 \times A} \frac{c_i(x,a)}{q_x(a)} \mathbf{P}_\gamma^{\sigma_\pi^M}(X_n \in dx, A_{n+1} \in da),$$

where the first equality is by the definition of S_3, and the last equality is by (6.28). Combining these observations, we see from (6.30) that for each $i = 0, 1, \ldots, N$,

$$\begin{aligned} E_\gamma^{\phi_\pi}\left[\int_0^\infty \int_A c_i(\xi_t, a)\phi_\pi(da|\xi_t)\, dt\right] &= \sum_{n=0}^\infty \int_{S\times A} \frac{c_i(x,a)}{q_x(a)} \mathbf{P}_\gamma^{\sigma_\pi^M}(X_n \in dx, A_{n+1} \in da) \\ &= \mathbf{E}_\gamma^{\sigma_\pi^M}\left[\sum_{n=0}^\infty \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right]. \end{aligned} \tag{6.31}$$

On the other hand, one can apply Theorem 3.3 of Dufour *et al.* [10] and the arguments in the proof of Lemma 4.2 for the existence of a stationary policy σ_π^S for the DTMDP model $\{S_\infty, A, p, \gamma\}$ satisfying that for each $i = 0, 1, \ldots, N$,

$$\mathbf{E}_\gamma^{\sigma_\pi^S}\left[\sum_{n=0}^\infty \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] \le \mathbf{E}_\gamma^{\sigma_\pi^M}\left[\sum_{n=0}^\infty \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right].$$

This and (6.31) thus prove the statement. □

Lemma 6.1. *Suppose Condition 4.1 is satisfied. Consider a stationary policy σ^S for the DTMDP model $\{S_\infty, A, p, \gamma\}$, which satisfies*

$$\sigma^S(da|x) = \delta_{f^*(x)}(da), \qquad \forall x \in S_2,$$

and is optimal and with a finite value for problem (3.3). Here the transition probability $p(dy|x,a)$ is given by (3.4) and (3.5). Then there is a stationary policy π^S for the CTMDP problem (2.2) satisfying for each $i = 0, 1, \ldots, N$,

$$\mathbf{E}_\gamma^{\sigma^S}\left[\sum_{n=0}^\infty \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right] = E_\gamma^{\pi^S}\left[\int_0^\infty \int_A c_i(\xi_t, a)\pi^S(da|\xi_t)\, dt\right].$$

Proof. Since σ^S is feasible with a finite value for problem (3.3), it is easy to see that $\sum_{n=0}^\infty \mathbf{P}_\gamma^{\sigma^S}(X_n \in \hat{S}_1) = 0$ so that, if necessary, we can modify the definition of the policy σ^S by putting

$$\sigma^S(da|x) = \delta_\Delta(da), \qquad \forall x \in \hat{S}_1,$$

with $\Delta \in A$ being an arbitrarily fixed point; the resulting policy is still optimal with a finite value for problem (3.3) and with the same performance vector as of the original policy.

Note also that $\sigma^S(B(x)|x) = 0$ for each $x \in S_1 \setminus \hat{S}_1$. For this reason, we can legitimately define the following stationary policy π^S for the CTMDP model;

$$\pi^S(da|x) = \frac{\frac{1}{q_x(a)}\sigma^S(da|x)}{\int_A \frac{1}{q_x(a)}\sigma^S(da|x)}$$

for each $x \in S \setminus (S_2 \cup \hat{S}_1)$,

$$\pi^S(da|x) = \delta_\Delta(da),$$

for each $x \in \hat{S}_1$, and

$$\pi^S(da|x) = \delta_{f^*(x)}(da)$$

for each $x \in S_2$. The discrete-time Markov chain $\{X_n\}$ under $\mathbf{P}_\gamma^{\sigma^S}$ can be regarded as the embedded chain of the pure jump time-homogeneous Markov process $\{\xi_t\}$ under $P_\gamma^{\pi^S}$; see [16]. Indeed, it holds on $\mathcal{B}(S)$ that, for each $x \in S \setminus (S_2 \cup \hat{S}_1)$,

$$\int_A p(dy|x,a)\sigma^S(da|x) = \int_A \frac{\tilde{q}(dy|x,a)}{q_x(a)}\sigma(da|x) = \frac{\int_A \tilde{q}(dy|x,a)\pi^S(da|x)}{\int_A q_x(a)\pi^S(da|x)};$$

for each $x \in S_2$,

$$\begin{aligned}\int_A p(dy|x,a)\sigma^S(da|x) &= \int_A \frac{\tilde{q}(dy|x,a)}{q_x(a)}\sigma^S(da|x) = \frac{\tilde{q}(dy|x,f^*(x))}{q_x(f^*(x))} \\ &= \frac{\int_A \tilde{q}(dy|x,a)\pi^S(da|x)}{\int_A q_x(a)\pi^S(da|x)} = 0;\end{aligned}$$

and for each $x \in \hat{S}_1$,

$$\int_A p(dy|x,a)\sigma^S(da|x) = \int_A \frac{\tilde{q}(dy|x,a)}{q_x(a)}\sigma^S(da|x) = \frac{\tilde{q}(dy|x,\Delta)}{q_x(\Delta)} = \frac{\int_A \tilde{q}(dy|x,a)\pi^S(da|x)}{\int_A q_x(a)\pi^S(da|x)}.$$

Furthermore, it is easy to verify that for each $i = 0, 1, \ldots, N$, given the current state $x \in S$, the (conditional) expected total cost during the current sojourn time of ξ_t under $P_\gamma^{\pi^S}$ is given by $\int_A \frac{c_i(x,a)}{q_x(a)}\sigma^S(da|x)$, which is the same as the (conditional) expected one-step cost for the discrete-time Markov chain $\{X_n\}$ under $\mathbf{P}_\gamma^{\sigma^S}$. The statement of this lemma now follows. □

Condition 6.1. *For problem* (2.2)*, there exists a feasible policy with a finite value.*

Theorem 6.2. *Suppose Condition* 4.1 *and Condition* 6.1 *are satisfied. Then for the CTMDP problem* (2.2)*, there is a stationary optimal policy* π*.*

Proof. It is clear that for the CTMDP problem (2.2), one can be restricted to the class of feasible policies π with a finite value and satisfying (6.1); there exists at least one such policy under Condition 6.1. It also holds that for the DTMDP problem (3.3), if the stationary policy σ_1^S for the DTMDP model $\{S_\infty, A, p, \gamma\}$ is optimal, then the stationary policy σ^S for the DTMDP model $\{S_\infty, A, p, \gamma\}$ defined by $\sigma^S(da|x) = \sigma_1^S(da|x)$ for each $x \in S \setminus S_2$, and $\sigma^S(da|x) = \delta_{f^*(x)}(da)$ for each $x \in S_2$ is also optimal with a finite value for problem (3.3). Now the statement is a consequence of Lemma 4.2, Lemma 6.1, Corollary 6.2, and Theorem 4.1 of [10]. □

Remark 6.1. Suppose Condition 4.1 and Condition 6.1 are satisfied. Theorem 4.1 of Dufour *et al.* [10], Lemma 4.2, Lemma 6.1 and Corollary 6.2 justify the reduction of problem (2.2) for the CTMDP model $\{S, A, q, \gamma\}$ to problem (3.3) for the DTMDP model $\{S_\infty, A, p, \gamma\}$ as well as to problem (4.4) for the DTMDP model $\{S, A, \tilde{p}, \gamma\}$; once the stationary optimal policy for the DTMDP problem (3.3) or for the DTMDP problem (4.4), which exists, is obtained, an optimal stationary policy for the CTMDP problem (2.2) can be automatically constructed based on it in principle, and the three problems have the same value.

Remark 6.2. As was rightly noted in [13], if the transition rates $q_x(a)$ are separated from zero, then one can show that for each policy π for the CTMDP, there is a policy σ for the DTMDP $\{S_\infty, A, p, \gamma\}$ such that

$$E_\gamma^\pi\left[\int_0^\infty \int_A c_i(\xi_t, a)\pi(da|\omega, t)\,dt\right] = \mathbf{E}_\gamma^\sigma\left[\sum_{n=0}^\infty \frac{c_i(X_n, A_{n+1})}{q_{X_n}(A_{n+1})}\right]$$

and vice versa, for each $i = 0, 1, \ldots, N$. The argument is essentially the same as for the discounted case, and the reduction is possible without further conditions. However, the objective of the present paper is to consider the more delicate and nontrivial case, that is, when the transition rates are not necessarily separated from zero.

7. Conclusion

To sum up, for the constrained total undiscounted optimal control problem for a CTMDP in Borel state and action spaces, under the compactness and continuity conditions, we showed the existence of an optimal stationary policy out of the class of general nonstationary ones. In the process, we justified the reduction of the CTMDP model to a DTMDP model. Several properties about the occupancy and occupation measures were obtained, too.

We mention that compared to discounted models, the total undiscounted criterion is significantly more challenging for studies. For DTMDP models, often the studies of this criterion are facilitated with some absorbing assumptions; see [2] and [17]. The absorbing assumption allows one to focus on the restriction of the occupation measure to a subset of the state space, where it is finite. The difficulty in dealing with undiscounted total criterion is the infiniteness of the occupation measure. In this connection, let us mention that DTMDP models with infinite occupation measures were studied in a recent series of papers [10–12].

Acknowledgements

This research is partially supported by NSFC and GDUPS. We thank the Editor and the three referees for their very helpful remarks. Especially two referees pointed out the flaw of a condition imposed in the earlier version of this paper, and the present Condition 4.1(b) is suggested by one referee.

References

[1] Aliprantis, C. and Border, K. (2007). *Infinite Dimensional Analysis*. New York: Springer.
[2] Altman, E. (1999). *Constrained Markov Decision Processes*. *Stochastic Modeling*. Boca Raton, FL: Chapman & Hall/CRC. MR1703380
[3] Bäuerle, N. and Rieder, U. (2009). MDP algorithms for portfolio optimization problems in pure jump markets. *Finance Stoch*. **13** 591–611. MR2519845
[4] Bäuerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Heidelberg: Springer. MR2808878
[5] Bertsekas, D.P. and Shreve, S.E. (1978). *Stochastic Optimal Control*: *The Discrete Time Case*. New York: Academic Press. MR0511544
[6] Borkar, V.S. (2002). Convex analytic methods in Markov decision processes. In *Handbook of Markov Decision Processes*. *Internat*. *Ser*. *Oper*. *Res*. *Management Sci*. **40** 347–375. Boston, MA: Kluwer Academic. MR1887208
[7] Costa, O.L.d.V. and Dufour, F. (2013). *Continuous Average Control of Piecewise Deterministic Markov Processes*. *Springer Briefs in Mathematics*. New York: Springer. MR3059228
[8] Davis, M.H.A. (1993). *Markov Models and Optimization*. *Monographs on Statistics and Applied Probability* **49**. London: Chapman & Hall. MR1283589
[9] Derman, C. and Strauch, R.E. (1966). A note on memoryless rules for controlling sequential control processes. *Ann*. *Math*. *Stat*. **37** 276–278. MR0184778
[10] Dufour, F., Horiguchi, M. and Piunovskiy, A.B. (2012). The expected total cost criterion for Markov decision processes under constraints: A convex analytic approach. *Adv. in Appl*. *Probab*. **44** 774–793. MR3024609
[11] Dufour, F. and Piunovskiy, A.B. (2010). Multiobjective stopping problem for discrete-time Markov processes: Convex analytic approach. *J*. *Appl*. *Probab*. **47** 947–966. MR2752898
[12] Dufour, F. and Piunovskiy, A.B. (2013). The expected total cost criterion for Markov decision processes under constraints. *Adv. in Appl*. *Probab*. **45** 837–859. MR3102474
[13] Feinberg, E.A. (2004). Continuous time discounted jump Markov decision processes: A discrete-event approach. *Math*. *Oper*. *Res*. **29** 492–524. MR2082616
[14] Feinberg, E.A. (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*. *Systems Control Found*. *Appl*. 77–97. Springer, New York. Birkhäuser. MR2961380
[15] Feinberg, E.A., Kasyanov, P.O. and Zadoianchuk, N.V. (2013). Berge's theorem for noncompact image sets. *J*. *Math*. *Anal*. *Appl*. **397** 255–259. MR2968988
[16] Feinberg, E.A., Mandava, M. and Shiryaev, A.N. (2014). On solutions of Kolmogorov's equations for nonhomogeneous jump Markov processes. *J*. *Math*. *Anal*. *Appl*. **411** 261–270. MR3118483
[17] Feinberg, E.A. and Rothblum, U.G. (2012). Splitting randomized stationary policies in total-reward Markov decision processes. *Math*. *Oper*. *Res*. **37** 129–153. MR2891151
[18] Forwick, L., Schäl, M. and Schmitz, M. (2004). Piecewise deterministic Markov control processes with feedback controls and unbounded costs. *Acta Appl*. *Math*. **82** 239–267. MR2069521

[19] Guo, X. (2007). Continuous-time Markov decision processes with discounted rewards: The case of Polish spaces. *Math. Oper. Res.* **32** 73–87. MR2292498
[20] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes*: *Theory and Applications*. *Stochastic Modelling and Applied Probability* **62**. Berlin: Springer. MR2554588
[21] Guo, X. and Piunovskiy, A. (2011). Discounted continuous-time Markov decision processes with constraints: Unbounded transition and loss rates. *Math. Oper. Res.* **36** 105–132. MR2799395
[22] Guo, X., Vykertas, M. and Zhang, Y. (2013). Absorbing continuous-time Markov decision processes with total cost criteria. *Adv. in Appl. Probab.* **45** 490–519. MR3102460
[23] Hernández-Lerma, O. and Lasserre, J.B. (1996). *Discrete-Time Markov Control Processes*. New York: Springer. MR1363487
[24] Hernández-Lerma, O. and Lasserre, J.B. (2000). Fatou's lemma and Lebesgue's convergence theorem for measures. *J. Appl. Math. Stoch. Anal.* **13** 137–146. MR1768500
[25] Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon–Nikodým derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* **31** 235–253. MR0380978
[26] Kitaev, M.Yu. (1985). Semi-Markov and jump Markov controllable models. Average cost criterion. *Theory Probab. Appl.* **30** 272–288.
[27] Kitaev, M.Yu. and Rykov, V.V. (1995). *Controlled Queueing Systems*. Boca Raton, FL: CRC Press. MR1413045
[28] Lippman, S.A. (1975). Applying a new device in the optimization of exponential queuing systems. *Oper. Res.* **23** 687–710. MR0443125
[29] Piunovskiĭ, A. (1998). A controlled discounted jump model under constraints. *Theory Probab. Appl.* **42** 51–71.
[30] 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. MR2837510
[31] Piunovskiy, A. and Zhang, Y. (2012). The transformation method for continuous-time Markov decision processes. *J. Optim. Theory Appl.* **154** 691–712. MR2945241
[32] Piunovskiy, A.B. (1997). *Optimal Control of Random Sequences in Problems with Constraints*. Dordrecht: Kluwer Academic. MR1472738
[33] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games*. London: Imperial College Press. MR2977542
[34] Puterman, M.L. (1994). *Markov Decision Processes*: *Discrete Stochastic Dynamic Programming*. New York: Wiley. MR1270015
[35] Schäl, M. (1975). Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. *Z. Wahrsch. Verw. Gebiete* **32** 179–196. MR0378841
[36] Serfozo, R.F. (1979). An equivalence between continuous and discrete time Markov decision processes. *Oper. Res.* **27** 616–620. MR0533923
[37] Yushkevich, A. (1980). On reducing a jump controllable Markov model to a model with discrete time. *Theory Probab. Appl.* **25** 58–68.

Received April 2015 and revised September 2015