Adv. Appl. Prob. **47,** 1064–1087 (2015)
Printed in Northern Ireland

FINITE-HORIZON OPTIMALITY FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES WITH UNBOUNDED TRANSITION RATES

XIANPING GUO,* **
XIANGXIANG HUANG * *** AND
YONGHUI HUANG,* **** *Sun Yat-Sen University*

Abstract

In this paper we focus on the *finite-horizon* optimality for denumerable continuous-time Markov decision processes, in which the transition and reward/cost rates are allowed to be unbounded, and the optimality is over the class of all randomized *history-dependent* policies. Under mild reasonable conditions, we first establish the existence of a solution to the finite-horizon optimality equation by designing a technique of approximations from the bounded transition rates to unbounded ones. Then we prove the existence of $\varepsilon(\geq 0)$-optimal Markov policies and verify that the value function is the unique solution to the optimality equation by establishing the *analog* of the Itô–Dynkin formula. Finally, we provide an example in which the transition rates and the value function are all *unbounded* and, thus, obtain solutions to some of the unsolved problems by Yushkevich (1978).

Keywords: Continuous-time Markov decision process; finite-horizon criterion; optimal Markov policy; randomized history-dependent policy; unbounded transition rate

2010 Mathematics Subject Classification: Primary 90C40
Secondary 93E20; 60J27

1. Introduction

Continuous-time Markov decision processes (CTMDPs) have been widely studied due to their rich applications in telecommunication, queueing systems, population processes, epidemiology, and so on; see, e.g. the survey [11], the monographs [8], [24], the recent works [9], [12], [20], [21], and [25], and the extensive references therein. As is well known, the commonly used optimality criteria in CTMDPs are the expected *discounted, average,* and the *finite-horizon*. The former two criteria are on the infinite (time-) horizon case, and have been well studied; see, [3], [4], [7], [8], [9], [11], [16], [20], [21], [23], [24], and [26] for the infinite-horizon expected discounted criterion and [8], [10], [11], [12], [17], and [24] for the long run expected average criterion. In this paper we focus on the *finite-horizon* criterion for CTMDPs, thus we shall not pinpoint the earlier literature on the average and discounted CTMDPs with an infinite horizon, and give emphasis to those on finite-horizon CTMDPs. In fact, only a few works address the finite-horizon CTMDPs. Miller [19] studied the finite-horizon finite-state CTMDPs with finite actions and within the class of Markov policies, and gave a necessary and sufficient condition

Received 1 October 2013; revision received 22 October 2014.
* Postal address: School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, 510275, P. R. China.
** Email address: mcsgxp@mail.sysu.edu.cn
*** Email address: hxiangx3@163.com
**** Email address: hyongh5@mail.sysu.edu.cn

for the existence of a piecewise constant optimal policy. Yushkevich [27] studied the finite-horizon denumerable-state CTMDPs with uniformly bounded transition rates and within the class of all deterministic history-dependent policies, and established the existence of an optimal Markov policy for the case of bounded rewards. He suggested that it is an *unsolved problem* to do away with the required boundedness of v_t in [27, Theorem 5.1, p. 216 and Theorem 5.2, p. 234], where v_t denotes the value function. Pliska [22] studied the finite-horizon general-state CTMDPs with uniformly bounded transition rates and within Markov policies, and showed the existence of an optimal Markov policy. Baüerle and Rieder [1] considered finite-horizon denumerable-state CTMDPs with bounded transition rates and within the class of deterministic Markov policies. They transformed the finite-horizon CTMDPs to *equivalent infinite-horizon* discrete-time Markov decision processes and thereby established the optimality equation and the existence of an optimal Markov policy using the existing theory on discrete-time Markov decision processes. Recently, Ghosh and Saha [6] investigated the finite-horizon general-state CTMDPs with uniformly bounded transition rates and within Markov policies established the existence of a unique solution of the optimality equation by the Banach fixed point theorem, and also proved the existence of an optimal Markov policy using the Itô–Dynkin formula. Note that all existing works [1], [6], [19], [22], and [27] (on finite-horizon CTMDPs) are limited to the case of uniformly bounded transition/reward rates and the history-independent policies in [1], [6], [19], and [22]. This boundedness requirement, however, imposes some restrictions in applications, for instance in queueing control and population processes, where the transition and reward/cost rates are usually unbounded [8], [11], [20], [21], and [24]. Hence, it is natural to consider finite-horizon CTMDPs with unbounded transition and reward/cost rates and extend the main results in [1], [6], [19], [22], and [27] to the case of randomized history-dependent policies and unbounded transition rates. Furthermore, it is desirable to find some solutions to the aforementioned unsolved problems in [27].

As indicated above, the finite-horizon CTMDPs with unbounded transition rates and within randomized history-dependent policies have *not* been studied yet, and they will be considered in this paper. More precisely, we will deal with the CTMDPs having the following features:

1. the transition rates may be *unbounded* and depend on time;
2. the reward/cost rates may be *time-dependent* and *unbounded from both above and below*;
3. the states are denumerable and the action space is a Borel space;
4. the policies can be randomized and *history-dependent*;
5. the optimality criterion is the *finite-horizon* expected rewards/costs.

First, under mild conditions slightly weaker than those in [8], [9], [11], [12], [20], [21], [24], and [25] on infinite-horizon CTMDPs, from the analog of the forward Kolmogorov equation developed recently in [9], [12], we establish the *analog* of the Itô–Dynkin formula for the underlying processes induced by the transition rates and *randomized history-dependent* policies (see Theorem 3.1 below). This result is a natural extension of the well-known Itô–Dynkin formula of a jump Markov process in [6], [8], [11], [19], and [22] to the case of a 'non-Markov' process.

Second, under suitable conditions as in [7]–[9], [11], [12], [20], [21], [24], and [25] on CTMDPs with infinite horizon, we prove the existence of a solution to the optimality equation for the finite-horizon CTMDPs in two steps. The first step is to prove the existence of a solution to the optimality equation for the case of bounded transition rates but unbounded rewards/costs

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

by the Banach fixed point theorem; see Proposition 4.1. This step is a generalization of that in Ghosh and Saha [6] for bounded costs to the case of unbounded rewards/costs. The second step is as follows. By designing a technique of approximations from bounded transition rates to unbounded ones, we further establish the existence of a solution to the optimality equation for the case of unbounded transition and reward/cost rates using the Ascoli theorem; see Theorem 4.1. The second step is *new* and *crucial* for the finite-horizon CTMDPs with unbounded transition rates.

Third, using the *analog* of the Itô–Dynkin formula developed here, from the optimality equation for the finite-horizon CTMDPs we prove the existence of $\varepsilon(\geq 0)$-optimal Markov policies, and also show that the value function of the finite-horizon CTMDPs is the unique solution to the optimality equation; see Proposition 4.1 and Theorem 4.1. All arguments here are direct, need no result from the existing theory on discrete-time Markov decision processes and, thus, are different from those in [1] and [27].

Finally, to illustrate our main results, we present an example in which our conditions are satisfied and the value function is unbounded. Moreover, the exact forms of an optimal Markov policy and the *unbounded* value function are obtained for two special cases of the example. This implies that the required boundedness of v_t (the value function) in [27, Theorem 5.1, p. 216 and Theorem 5.2, p. 234] can be done away with and, thus, some of the unsolved problems by Yushkevich [27] will have been solved; see Remark 5.1 and Proposition 5.1. Also, the conditions in this paper are slightly weaker than those in [7]–[9], [11], [12], [21], [20], [24], and [25] (see Remarks 3.1 and 3.2 below) and, thus, all existing examples therein satisfy the conditions in this paper. Furthermore, it is easy to provide examples which can verify all conditions in this paper but do not satisfy some of conditions in [7]–[9], [11], [12], [21], [20], [24], and [25].

The rest of the paper is organized as follows. In Section 2 we introduce the optimality problem for the finite-horizon CTMDPs. The main results are presented in Section 4 after giving technical preliminaries in Section 3, and are illustrated with an example in Section 5.

2. The optimal control problems

Notation. For any Borel space X endowed with the Borel σ-algebra $\mathcal{B}(X)$, we denote by $\mathcal{U}(X)$ the universal σ-algebra on X, i.e. $\mathcal{U}(X) := \bigcap_{p\in\mathbb{P}(X)}\mathcal{B}_p(X)$, where $\mathbb{P}(X)$ represents the set of all probability measures on X and $\mathcal{B}_p(X)$ is the completion of $\mathcal{B}(X)$ with respect to $p \in \mathbb{P}(X)$. To discern the 'measurability' we will say 'Borel measurable' or 'universally measurable' in the following. The nonhomogeneous model of CTMDPs is a six-tuple

$$\{S, A, A(t,i)(t \geq 0,\ i \in S), r(t,i,a), q(j \mid t,i,a), g(t,i)\} \tag{2.1}$$

consisting of the following elements:

(i) a denumerable set S, called the state space, whose elements are referred to as states of a system;

(ii) a Borel space A, called the action space, whose elements are referred to as actions (or decisions) of a decision maker (or controller);

(iii) a family $\{A(t,i),\ t \geq 0,\ i \in S\}$ of nonempty subsets $A(t,i)$ of A, where each $A(t,i)$ denotes the set of actions available to a controller when the system is in state $i \in S$ at time t, and it is assumed to be Borel measurable, i.e. $A(t,i) \in \mathcal{B}(A)$ for all $t \geq 0$ and $i \in S$;

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

(iv) a Borel measurable function $r(t, i, a)$ on $\mathbb{K}$, called the reward rates, where $\mathbb{K} := \{(t, i, a) \mid t \in [0, \infty),\ i \in S,\ a \in A(t, i)\}$;

(v) a real-valued function $g(t, i)$ on $[0, \infty) \times S$, called the terminal reward at time t (*As $r(t, i, a)$ and $g(t, i)$ are allowed to take positive and negative values they can be interpreted as costs rather than 'rewards' only.*);

(vi) transition rates $q(j \mid t, i, a)$, a Borel measurable signed kernel on S given $\mathbb{K}$, taking nonnegative values for all $j \neq i$ with $j, i \in S$, being conservative in the sense of $q(S \mid t, i, a) \equiv 0$ and stable in that of

$$q^*(i) := \sup_{t \geq 0, a \in A(t,i)} q(t, i, a) < \infty \quad \text{for all } i \in S, \tag{2.2}$$

where $q(t, i, a) := -q(i \mid t, i, a) \geq 0$ for all $(t, i, a) \in \mathbb{K}$.

The model (2.1) is called homogeneous if the data in (2.1) are independent of time t.

Next, we provide an informal description of the evolution of CTMDPs with the model (2.1). Roughly speaking, a continuous-time Markov decision process evolves as follows. A controller observes states of a system continuously in time. If the system remains at state i at time t, he/she chooses an action $a \in A(t, i)$ (possibly dependent on histories) according to some given policy, as a consequence of which, the following happens:

(i) an immediate reward/cost takes place at the rate $r(t, i, a)$;

(ii) after a random sojourn time (i.e. the holding time at state i), the system jumps to a new state j with the transition probability $q(j \mid t, i, a)/q(t, i, a)$. The nonhomogeneous exponential distribution of sojourn times is $(1 - \exp(-\int_0^t q(s, i, a)\,\mathrm{d}s))$ determined by the transition rates $q(j \mid t, i, a)$.

To formalize what is described above, below we describe the construction of CTMDPs under possibly randomized history-dependent policies. To this end, we introduce some notation. Let $S_\Delta := S \cup \{\Delta\}$ (with some $\Delta \notin S$), $\Omega^0 := (S \times (0, \infty))^\infty$, $\Omega := \Omega^0 \cup \{(i_0, \theta_1, i_1, \ldots, \theta_k, i_k, \infty, \Delta, \infty, \ldots) \mid i_0 \in S, i_l \in S,\ \theta_l \in (0, \infty) \text{ for each } 1 \leq l \leq k,\ k \geq 1\}$ and let $\mathcal{F}$ be the universal σ-algebra on Ω. Then we obtain the measurable space $(\Omega, \mathcal{F})$. For each $k \geq 0$, $e := (i_0, \theta_1, i_1, \ldots, \theta_k, i_k, \ldots) \in \Omega$, let $h_k(e) := (i_0, \theta_1, i_1, \ldots, \theta_k, i_k)$ denote the k-component internal history, and define

$$T_0(e) := 0, \qquad T_{k+1}(e) := \theta_1 + \theta_2 + \cdots + \theta_{k+1}, \qquad X_k(e) := i_k.$$

In what follows, the argument e is always omitted. Let $T_\infty := \lim_{k \to \infty} T_k$, and define the state process $\{x_t\}$ by

$$x_t := \sum_{k \geq 0} \mathbf{1}_{\{T_k \leq t < T_{k+1}\}}\, i_k + \Delta\, \mathbf{1}_{\{t \geq T_\infty\}} \quad \text{for } t \geq 0. \tag{2.3}$$

Here and below, $\mathbf{1}_E$ stands for the indicator function on any set E.

From (2.3), we see that T_k ($k \geq 1$) denotes the kth jump moment of $\{x_t\}$, i_{k-1} is the state of the process on $[T_{k-1}, T_k)$, and θ_k plays the role of sojourn times at state i_{k-1}. We do not intend to consider the controlled process after moment T_∞ and, thus, view it to be absorbed in the cemetery state $\Delta \notin S$. Hence, we write $q(\cdot \mid t, \Delta, a_\Delta) :\equiv 0$, $r(t, \Delta, a_\Delta) :\equiv 0$, $A(t, \Delta) := \{a_\Delta\}$, and $A_\Delta := A \cup \{a_\Delta\}$, where a_Δ is an isolated point.

Take the right-continuous family of σ-algebras $\{\mathcal{F}_t\}_{t \geq 0}$ as the internal history of the marked point process $\{T_k, X_k, k \geq 0\}$, i.e. $\mathcal{F}_t := \sigma(T_m \leq s, X_m = i, i \in S, s \leq t, m \geq 0)$. Let $\mathcal{P}$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

be the universal σ-algebra of predictable sets on $\Omega \times [0,\infty)$ related to $\{\mathcal{F}_t\}_{t\geq 0}$, i.e. $\mathcal{P} := \sigma(\{\Gamma \times \{0\}, \Gamma \in \mathcal{F}_0\} \cup \{\Gamma \times (s,\infty), \Gamma \in \mathcal{F}_{s-}, s > 0\})$, where $\mathcal{F}_{s-} := \vee_{t<s}\mathcal{F}_t$; see [18, Chapter 4] for details. A real-valued function on $\Omega \times [0,\infty)$ is called predictable if it is measurable with respect to $\mathcal{P}$.

To precisely define the optimality criterion, we need to introduce the concept of a policy, which is a generalization of the policies (on Borel measurability) in [9], [12], [18], [20], and [21] to the universal measurability.

Definition 2.1. A *randomized history-dependent policy* is a $\mathcal{P}$-measurable transition probability $\pi(\mathrm{d}a \mid e,t)$ from $\Omega \times [0,\infty)$ onto A_Δ, which is concentrated on $A(t,x_{t-})$, where $x_{t-} = \lim_{s\uparrow t} x_s$. A policy $\pi(\mathrm{d}a \mid e,t)$ is called randomized Markov if it has the form $\pi(\mathrm{d}a \mid x_{t-},t)$, which is denoted by $\pi_{\mathrm{t}}(\mathrm{d}a \mid \cdot)$ for informational implication. A randomized Markov policy $\pi_{\mathrm{t}}(\mathrm{d}a \mid \cdot)$ is called a (deterministic) Markov policy whenever there exists an A-valued and universally measurable function $f(t,i)$ on $[0,\infty) \times S$ such that $\pi_{\mathrm{t}}(\mathrm{d}a \mid i)$ is a Dirac measure concentrated at $f(t,i)$. Such a Markov policy will be denoted by f for simplicity.

We denote by Π the set of all randomized history-dependent policies, by $\Pi^{\mathrm{r}}_{\mathrm{m}}$ the set of all randomized Markov policies, and by $\Pi^{\mathrm{d}}_{\mathrm{m}}$ the set of all deterministic Markov policies.

Due to the predictability of a policy, from [18, Theorems 4.13 and 4.19 or Equation (4.38)] it can be seen that each policy $\pi(\mathrm{d}a \mid e,t)$ can be characterized by the following expression:

$$\pi(\mathrm{d}a \mid e,t) = \mathbf{1}_{\{t=0\}}\pi^0(\mathrm{d}a \mid i_0,0) + \sum_{k\geq 0}\mathbf{1}_{\{T_k<t\leq T_{k+1}\}}\pi^k(\mathrm{d}a \mid i_0,\theta_1,i_1,\ldots,\theta_k,i_k,t-T_k) + \mathbf{1}_{\{t\geq T_\infty\}}\delta_{a_\Delta}(\mathrm{d}a), \tag{2.4}$$

where $\pi^0(\mathrm{d}a \mid i_0,0)$ is a stochastic kernel on A given S, $\pi^k(k\geq 1)$ are stochastic kernels on A given $(S\times(0,\infty))^{k+1}$, and $\delta_{a_\Delta}(\mathrm{d}a)$ denotes the Dirac measure at the point a_Δ.

Evidently, for any policy $\pi \in \Pi$, the random measure

$$m^\pi(j \mid e,t)\,\mathrm{d}t := \int_A q(j \mid t,x_{t-},a)\pi(\mathrm{d}a \mid e,t)\,\mathbf{1}_{\{j\neq x_{t-}\}}\,\mathrm{d}t \tag{2.5}$$

is predictable. Note that $m^\pi(j \mid e,t)$ in (2.5) defines the jumps intensity of the process $\{x_t\}$, which together with (2.4) gives the following representation:

$$m^\pi(j \mid e,t) = \mathbf{1}_{\{t=0\}}m^\pi_0(j \mid i_0,0) + \sum_{k\geq 0}\mathbf{1}_{\{T_k<t\leq T_{k+1}\}}m^\pi_k(j \mid i_0,\theta_1,i_1,\ldots,\theta_k,i_k,t-T_k), \tag{2.6}$$

where $m^\pi_k(j \mid i_0,\theta_1,i_1,\ldots,\theta_k,i_k,t-T_k) := \int_A q(j \mid t,i_k,a)\pi^k(\mathrm{d}a \mid i_0,\theta_1,\ldots,\theta_k,i_k,t-T_k)\mathbf{1}_{\{j\neq i_k\}}$ for $T_k < t \leq T_{k+1}$, $m^\pi_0(j \mid i_0,0) := \int_A q(j \mid 0,i_0,a)\pi^0(\mathrm{d}a \mid i_0,0)\mathbf{1}_{\{j\neq i_0\}}$; see [15] for details.

For any initial distribution γ on S and policy $\pi \in \Pi$, let us recall the structure of the measure $\mathbb{P}^\pi_\gamma$ on the measurable space $(\Omega,\mathcal{F})$ given in [9], [12], [20], and [21]. Let $H_0 := S$ and $H_k := S\times((0,\infty]\times S_\Delta)^k$, $k=1,2,\ldots$. The measure $\mathbb{P}^\pi_\gamma$ on H_0 is given by $\mathbb{P}^\pi_\gamma(i) = \gamma(i)$ for all $i \in S$. Suppose that measure $\mathbb{P}^\pi_\gamma$ on H_k has been constructed. Actually, $\mathbb{P}^\pi_\gamma$ will be a measure on $(\Omega,\mathcal{F})$, but here we deal with its marginal projection onto the space of k-component

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

histories H_k. Then $\mathbb{P}_\gamma^\pi$ on H_{k+1} is determined by the following expressions:

$$\mathbb{P}_\gamma^\pi(\Gamma \times (\mathrm{d}t, j)) := \int_\Gamma \mathbb{P}_\gamma^\pi(\mathrm{d}h_k)\, \mathbf{1}_{\{\theta_k<\infty\}}\, m_k^\pi(j \mid h_k, t) \exp\left(-\int_0^t m_k^\pi(S \mid h_k, v)\,\mathrm{d}v\right)\mathrm{d}t,$$
$$\mathbb{P}_\gamma^\pi(\Gamma \times (\infty, \Delta)) := \int_\Gamma \mathbb{P}_\gamma^\pi(\mathrm{d}h_k)\left\{\mathbf{1}_{\{\theta_k=\infty\}} + \mathbf{1}_{\{\theta_k<\infty\}} \exp\left(-\int_0^\infty m_k^\pi(S \mid h_k, v)\,\mathrm{d}v\right)\right\}, \tag{2.7}$$

where $\Gamma \in \mathcal{U}(H_k)$ and $m_k^\pi(S \mid h_k, t) = \sum_{j\neq i_k} m_k^\pi(j \mid h_k, t)$.

According to the extension of the well-known Ionescu Tulcea theorem (see, e.g. [2, Proposition 7.45]), there exists a unique probability measure $\mathbb{P}_\gamma^\pi$ on $(\Omega, \mathcal{F})$ which has a projection onto H_k satisfying (2.7). Let $\mathbb{E}_\gamma^\pi$ be its corresponding expectation operator. In particular, $\mathbb{E}_\gamma^\pi$ and $\mathbb{P}_\gamma^\pi$ will be respectively written as $\mathbb{E}_i^\pi$ and $\mathbb{P}_i^\pi$ when γ is the Dirac measure located at state i in S.

Fix a constant $T > 0$, which denotes the finite horizon of the CTMDPs and is different from the variables T_k in (2.3) above. We now state the T-horizon optimality problem of the CTMDPs we are concerned with. For each policy $\pi \in \Pi$ and initial state $i \in S$, the expected T-horizon criterion $V_\pi(0, i)$ is defined by

$$V_\pi(0, i) := \mathbb{E}_i^\pi\left[\int_0^T \int_A r(t, x_t, a)\pi(\mathrm{d}a \mid e, t)\,\mathrm{d}t + g(T, x_T)\right],$$

provided that the integral is well defined. The T-horizon value function of the CTMDPs is

$$V^*(0, i) := \sup_{\pi\in\Pi} V_\pi(0, i) \quad \text{for } i \in S.$$

Note that the process $\{x_t, t \geq 0\}$ on $(\Omega, \mathcal{F}, \mathbb{P}_\gamma^\pi)$ may *not* be Markovian since the policy π can depend on histories. However, for each $\pi := \pi_{\mathrm{t}}(\mathrm{d}a \mid \cdot) \in \Pi_{\mathrm{m}}^{\mathrm{r}}$, it is well known that $\{x_t, t \geq 0\}$ is a jump Markov process; see, e.g. [5, Theorem 2.2]. Thus, for each $i \in S$ and $t \in [0, T]$, the following expressions are well defined (when the integral exists):

$$\mathbb{E}_{t,i}^\pi g(T, x_T) := \mathbb{E}_\gamma^\pi[g(T, x_T) \mid x_t = i],$$
$$\mathbb{E}_{t,i}^\pi\left[\int_t^T r(s, x_s, \pi_s)\,\mathrm{d}s + g(T, x_T)\right] := \mathbb{E}_\gamma^\pi\left[\int_t^T r(s, x_s, \pi_s)\,\mathrm{d}s + g(T, x_T) \mid x_t = i\right],$$

where $r(s, i, \pi_s) := \int_{A(s,i)} r(s, i, a)\pi_s(\mathrm{d}a \mid i)$.

The value of a policy $\pi \in \Pi_{\mathrm{m}}^{\mathrm{r}}$ from the horizon t to T, $V_\pi(t, i)$, is defined by

$$V_\pi(t, i) := \mathbb{E}_{t,i}^\pi\left[\int_t^T r(s, x_s, \pi_s)\,\mathrm{d}s + g(T, x_T)\right].$$

Let

$$V^*(t, i) := \sup_{\pi\in\Pi_{\mathrm{m}}^{\mathrm{r}}} V_\pi(t, i) \quad \text{for } (t, i) \in (0, T] \times S.$$

The function $V^*(t, i)$ ($t \in [0, T]$, $i \in S$) is called the value function of the finite-horizon CTMDPs from the horizon t to T.

Concerning the value function $V^*(t, i)$, we state the unsolved problems in Yushkevich [27, p. 216, p. 234].

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

> 'Unsolved problems. In analogy to the discrete time case it would be desirable to extend Theorems 4.1 and 4.2 to arbitrary summable models and in Theorems 5.1 and 5.2 to do away with the required boundedness of v_t.'

Note that v_t is the value function here.

Definition 2.2. For any given $\varepsilon \geq 0$, a policy $\pi^* \in \Pi$ is said to be ε-optimal if $V_{\pi^*}(0, i) \geq V^*(0, i) - \varepsilon$ for all $i \in S$. A 0-optimal policy is called an optimal policy.

The main goal of this paper is to provide conditions for the existence of ε-optimal Markov policies and also for the existence of solutions to the above unsolved problems [27, Theorem 5.1, p. 216 and Theorem 5.2, p. 234] for the finite-horizon CTMDPs.

3. Preliminaries

In this section we state some basic assumptions and preliminary facts that are needed to prove our main results. In particular, the *analog* of the Itô–Dynkin formula for the process $\{x_t, t \geq 0\}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P}^\pi_\gamma)$ associated with the unbounded transition rates and randomized history-dependent policies is derived.

Since the transition rates $q(j \mid t, i, a)$ and the reward function $r(t, i, a)$ may be unbounded, we need to establish the nonexplosion of $\{x_t,\ t \geq 0\}$ (i.e. $\mathbb{P}^\pi_i(T_\infty = \infty) = 1$ or $\mathbb{P}^\pi_i(x_t \in S) \equiv 1$) and the finiteness of the value function $V^*(t, i)$. To do so, we provide the following condition.

Assumption 3.1. *There exist a function $\omega \geq 1$ on S and constants $c > 0$, $b \geq 0$, and $M_1 > 0$ such that:*

(i) $\sum_{j \in S} q(j \mid t, i, a)\omega(j) \leq c\omega(i) + b$ for all $(t, i, a) \in \mathbb{K}$;

(ii) *there exists a sequence $\{S_m,\ m \geq 1\}$ of subsets of S such that $S_m \uparrow S$, $\sup_{i \in S_m} q^*(i) < \infty$, and $\lim_{m \to \infty} \inf_{j \notin S_m} \omega(j) = +\infty$, with $q^*(i)$ as in (2.2) and $\inf \varnothing := +\infty$;*

(iii) $|r(t, i, a)| \leq M_1\omega(i)$ *and* $|g(T, i)| \leq M_1\omega(i)$ *for each* $t \in [0, T]$, $i \in S$, *and* $a \in A(t, i)$.

Remark 3.1. (i) Assumption 3.1 is the extension of [9, Condition 3.1] and [21, Condition 1] for the homogeneous model to the nonhomogeneous case of $q(j \mid t, i, a)$ and $r(t, i, a)$. Thus, it is satisfied for the examples in [8], [9], [11], [20], [21], and [24]. Moreover, when the transition rates are bounded (i.e. $\sup_{i \in S} q^*(i) < \infty$) [3], [6], [19], [22], and [27], Assumptions 3.1(i) and 3.1(ii) are satisfied by taking $\omega(i) \equiv 1$ and $S_m \equiv S$. Assumption 3.1(iii) is required for the finiteness of $V^*(t, i)$.

(ii) If the number c in Assumption 3.1(i) is not positive, then Assumption 3.1(i) still holds when c is replaced with the positive number '$1 + |c|$'. Thus, for simplicity and convenience, we will assume that $c > 0$. However, the corresponding number is assumed to be negative in [8], [11], [12], [24], and [25] or less than the discount factor in [7]–[9], [20], and [21].

The following lemma from [9] and [21] establishes the nonexplosion of $\{x_t, t \geq 0\}$. We present it here for ease of reference.

Lemma 3.1. *Under Assumptions 3.1(i) and 3.1(ii) for each $\pi \in \Pi$, the following assertions hold:*

(i) $\mathbb{E}^\pi_i[\omega(x_t)] \leq e^{ct}[\omega(i) + b/c]$ *for each* $t \geq 0$ *and* $i \in S$;

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

(ii) $\mathbb{P}_i^\pi(x_t = j) = \delta_{ij} + \mathbb{E}_i^\pi[\int_0^t \int_A q(j \mid s, x_{s-}, a)\pi(\mathrm{d}a \mid e, s)\,\mathrm{d}s]$ *for each* $t \geq 0$ *and* $i, j \in S$, *where* δ_{ij} *is the Kronecker delta (i.e.* $\delta_{ii} = 1$ *for all* $i \in S$ *and* $\delta_{ij} = 0$ *for all* $i \neq j$*);*

(iii) $\sum_{j \in S} \mathbb{P}_i^\pi(x_t = j) = 1$ *for each* $t \geq 0$ *and* $i \in S$.

Proof. Replacing the $\Lambda(j \mid \omega, t)$ from the $[\int_A \pi(\mathrm{d}a \mid \omega, t) q(j \mid \xi_{t-}(\omega), a)\mathbf{1}_{\{j \neq \xi_{t-}\}}]\,\mathrm{d}t = \Lambda(j \mid \omega, t)\,\mathrm{d}t$ in [9, Equation (3)] (for the '$q(j \mid i, a)$') with $m^\pi(j \mid e, t) = \int_A q(j \mid t, x_{t-}, a)\pi(\mathrm{d}a \mid e, t)\mathbf{1}_{\{j \neq x_{t-}\}}$ in (2.5) for the time-dependent $q(j \mid t, i, a)$, we see that the representation of $m^\pi(j \mid e, t)$ in (2.6) is the same as that of $\Lambda(j \mid \omega, t)$ of [9, Equation (4)] with the obvious change of the symbols. Since the rest of the proof of [9, Theorem 3.1] depends only on $\Lambda^m(j \mid x_0, \theta_1, x_1, \ldots, \theta_m, x_m, t - T_m)$ in [9, Equation (4)], replacing the $\Lambda^m(j \mid x_0, \theta_1, x_1, \ldots, \theta_m, x_m, t - T_m)$ with $m_k^\pi(j \mid i_0, \theta_1, i_1, \ldots, \theta_k, i_k, t - T_k)$ in (2.6) and using the same arguments as in the proof of [9, Theorem 3.1], we see that this lemma holds.

The following result guarantees the finiteness of $V_\pi(s, i)(\pi \in \Pi_{\mathrm{m}}^{\mathrm{r}})$ and $V_\pi(0, i)(\pi \in \Pi)$.

Lemma 3.2. *Under Assumption 3.1, the following assertions hold:*

(i) $|V_\pi(0, i)| \leq (T + 1)M_1 \mathrm{e}^{cT}[\omega(i) + b/c]$ *for all* $i \in S$, $\pi \in \Pi$*;*

(ii) $|V_\pi(t, i)| \leq (T + 1)M_1 \mathrm{e}^{c(T-t)}[\omega(i) + b/c]$ *for all* $(t, i) \in [0, T] \times S$, $\pi \in \Pi_{\mathrm{m}}^{\mathrm{r}}$.

Proof of Lemma 3.2(i). For each $\pi \in \Pi$ and $i \in S$, by Lemma 3.1(i) and Assumption 3.1(iii), we have

$$\begin{aligned}
|V_\pi(0, i)| &= \left|\mathbb{E}_i^\pi\left[\int_0^T \int_A r(t, x_t, a)\pi(\mathrm{d}a \mid e, t)\,\mathrm{d}t + g(T, x_T)\right]\right| \\
&\leq \int_0^T M_1 \mathbb{E}_i^\pi \omega(x_t)\,\mathrm{d}t + M_1 \mathbb{E}_i^\pi \omega(x_T) \\
&\leq M_1 \int_0^T \left[\mathrm{e}^{ct}\omega(i) + \frac{b}{c}\mathrm{e}^{ct}\right]\mathrm{d}t + M_1\left[\mathrm{e}^{cT}\omega(i) + \frac{b}{c}\mathrm{e}^{cT}\right] \\
&\leq (T + 1)M_1 \mathrm{e}^{cT}\left[\omega(i) + \frac{b}{c}\right],
\end{aligned}$$

which implies Lemma 3.2(i).

Proof of Lemma 3.2(ii). The second statement follows from [8, Lemma 6.3].

Lemma 3.2 gives conditions for the finiteness of $V_\pi(s, i)(\pi \in \Pi_{\mathrm{m}}^{\mathrm{r}})$ and $V_\pi(0, i)(\pi \in \Pi)$. Lemma 3.1(i) gives the analog of the forward Kolmogorov equation, which will be used to derive the analog of the Itô–Dynkin formula for the process $\{x_t,\ t \geq 0\}$. To do so, it is necessary to introduce some further notation and conditions.

Assumption 3.2. *With* ω *as in Assumption 3.1, there exists a function* $\omega' \geq 1$ *on* S *and constants* $c' > 0, b' \geq 0$, *and* $M_2 > 0$ *such that*

$$q^*(i)\omega(i) \leq M_2\omega'(i), \qquad \sum_{j \in S} \omega'(j)q(j \mid t, i, a) \leq c'\omega'(i) + b' \quad \text{for all } (t, i, a) \in \mathbb{K},$$

where $q^*(i)$ *is as in (2.2).*

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Remark 3.2. Note that Assumption 3.2 is similar to [9, Condition 4.1] and [21, Condition 4], but compared with those in [9] and [21], the roles of ω and ω' here have been switched. In addition, Assumption 3.2 is for the nonhomogeneous case of $q(j \mid t, i, a)$, while [9, Condition 4.1] and [21, Condition 4] are for the homogeneous case, and neither the first part of [9, Condition 4.1(c)] nor [21, Condition 4(c)] is required here. Moreover, when the transition rates or the reward rates are bounded (i.e. $\sup_{(t,i,a)\in\mathbb{K}}|r(t,i,a)| < \infty$; see, for instance, [6]), Assumption 3.2 is not required. The role of Assumption 3.2 is for the finiteness of $\mathbb{E}_i^\pi[\omega(x_t)q^*(x_t)]$ for $t \geq 0$; see the assertions in (3.2) and (3.4) in proving Theorem 3.1 below.

Let $I := [0, T]$. Given any function $\bar{w} \geq 1$ on S, a real-valued function φ on $I \times S$ is called $\bar{w}$-bounded if the $\bar{w}$-weighted norm of φ, $\|\varphi\|_{\bar{w}} := \sup_{(t,i)\in I\times S}(|\varphi(t,i)|/\bar{w}(i))$, is finite. We denote by $B_{\bar{w}}(I \times S)$ the Banach space of all $\bar{w}$-bounded Borel measurable functions on $I \times S$. Moreover, a function $h(t, i)$ defined on $I \times S$ is called essentially $\bar{w}$-bounded if there exists a Borel subset Z of I such that $m(Z) = m(I)$ and $\|h\|_{\bar{w}}^{\text{es}} := \sup_{t\in Z, i\in S}(|h(t,i)|/\bar{w}(i)) < \infty$, where m is the Lebesgue measure on $[0, \infty)$.

For any $\varphi \in B_\omega(I \times S)$ and $i \in S$, if there is a set of Lebesgue-measure 0 (denoted by $L_\varphi(i) \subset I$) such that $\varphi(t, i)$ is differentiable in every $t \in L_\varphi^c(i) := I \setminus L_\varphi(i)$, we call $\varphi(t, i)$ differentiable almost everywhere, and denote by $\varphi'(t, i)$ the partial derivative of $\varphi(t, i)$. Since S is denumerable, $\bigcup_{i\in S} L_\varphi(i)$ is also a set of Lebesgue-measure 0 (i.e. $m(\bigcup_{i\in S} L_\varphi(i)) = 0$). From the point of view of Lebesgue-integration theory, functions that differ only on a set of Lebesgue-measure 0 are viewed as identified. Thus, since $\varphi'(t, i)$ is well defined at every $t \in I \setminus \bigcup_{i\in S} L_\varphi(i)$ and $i \in S$, in what follows, we can extend $\varphi'(t, i)$ on $(I \setminus \bigcup_{i\in S} L_\varphi(i)) \times S$ to a real-valued function on $I \times S$ by defining $\varphi'(t, i)$ to be 0 on $\bigcup_{i\in S} L_\varphi(i) \times S$, and such an extension of $\varphi'(t, i)$ makes no loss of generalization for the study on the criterion $V_\pi(t, i)$.

With ω and ω' as in Assumption 3.2, let $C_{\omega,\omega'}^{1,0}(I \times S) := \{\varphi \in B_\omega(I \times S)\colon \varphi(t, i)$ is absolutely continuous, and $\varphi'(t, i)$ is universally measurable in $t \in I$ (for each fixed $i \in S$) and essentially $(\omega + \omega')$-bounded on $I \times S\}$.

To prove the existence of an optimal policy, we need to introduce the Itô–Dynkin formula and derive its analog, which are given in the following theorem.

Theorem 3.1. *Suppose Assumptions 3.1(i), 3.1(ii), and 3.2 are satisfied. Then, for each* $\varphi \in C_{\omega,\omega'}^{1,0}(I \times S)$, *the following assertions hold.*

(i) *(The analog of the Itô–Dynkin formula.) For every* $i \in S$, $\pi \in \Pi$,

$$\mathbb{E}_i^\pi\left[\int_0^T\left(\varphi'(s, x_s) + \sum_{j\in S}\int_A \varphi(s, j)q(j \mid s, x_s, a)\pi(\mathrm{d}a \mid e, s)\right)\mathrm{d}s\right] = \mathbb{E}_i^\pi\varphi(T, x_T) - \varphi(0, i),$$

where $\{x_t,\ t \geq 0\}$ *may not be Markovian since the policy* π *may depend on histories.*

(ii) *(The Itô–Dynkin formula.) For each* $(t, i) \in I \times S$, $\pi = \pi_{\mathrm{t}}(\mathrm{d}a \mid \cdot) \in \Pi_{\mathrm{m}}^{\mathrm{r}}$,

$$\mathbb{E}_{t,i}^\pi\left[\int_t^T\left(\varphi'(s, x_s) + \sum_{j\in S}\varphi(s, j)q(j \mid s, x_s, \pi_s)\right)\mathrm{d}s\right] = \mathbb{E}_{t,i}^\pi\varphi(T, x_T) - \varphi(t, i),$$

where $q(j \mid s, k, \pi_s) := \int_{A(s,k)} q(j \mid s, k, a)\pi_s(\mathrm{d}a \mid k)$ *for all* $k, j \in S$, *and* $s \geq 0$.

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Proof of Theorem 3.1(i). Since $\varphi \in C^{1,0}_{\omega,\omega'}(I \times S)$, from the definition of $C^{1,0}_{\omega,\omega'}(I \times S)$ above, it follows that

$$|\varphi(s, j)| \le \|\varphi\|_\omega \omega(j), \qquad |\varphi'(s, j)| \le \|\varphi'\|^{\mathrm{es}}_{\omega+\omega'}(\omega(j) + \omega'(j)) \quad \text{for all } s \in L^c_\varphi(j),\ j \in S. \tag{3.1}$$

Thus, by Assumptions 3.1(i), 3.1(ii), and 3.2, we have

$$\begin{aligned}
&\sum_{j\in S}\int_A |q(j \mid s, k, a)\pi(\mathrm{d}a \mid e, s)\varphi(s, j)| \\
&\quad \le \|\varphi\|_\omega\left[\sum_{j\ne k}\int_A \omega(j)q(j \mid s, k, a)\pi(\mathrm{d}a \mid e, s) + \omega(k)q^*(k)\right] \\
&\quad \le \|\varphi\|_\omega\left[\sum_{j\in S}\int_A \omega(j)q(j \mid s, k, a)\pi(\mathrm{d}a \mid e, s) + 2\omega(k)q^*(k)\right] \\
&\quad \le \|\varphi\|_\omega[c\omega(k) + 2M_2\omega'(k) + b] \quad \text{for all } (s, k) \in I \times S.
\end{aligned} \tag{3.2}$$

Moreover, since $\bigcup_{j\in S} L_\varphi(j)$ is a set of Lebesgue-measure 0, by (3.1), we have

$$\int_0^T |\varphi'(s, x_s)|\,\mathrm{d}s \le \|\varphi'\|^{\mathrm{es}}_{\omega+\omega'}\int_0^T (\omega(x_s) + \omega'(x_s))\,\mathrm{d}s,$$

from which, together with Lemma 3.1(i) (with ω being replaced with $(\omega+\omega')$ here), we obtain

$$\mathbb{E}^\pi_i\left[\int_0^T |\varphi'(s, x_s)|\,\mathrm{d}s\right] \le \|\varphi'\|^{\mathrm{es}}_{\omega+\omega'} T\mathrm{e}^{(c+c')T}\left[\omega(i) + \omega'(i) + \frac{b+b'}{c+c'}\right] < \infty. \tag{3.3}$$

Thus, from (3.2) and Lemma 3.1(i), we obtain

$$\begin{aligned}
&\int_t^T \sum_{j\in S}\mathbb{E}^\pi_i \int_A |q(j \mid s, x_s, a)\pi(\mathrm{d}a \mid e, s)\varphi(s, j)|\,\mathrm{d}s \\
&\quad \le \|\varphi\|_\omega \int_0^T \mathbb{E}^\pi_i[cw(x_s) + b + 2M_2\omega'(x_s)]\,\mathrm{d}s \\
&\quad \le T\|\varphi\|_\omega\left[(c+b)\mathrm{e}^{cT}w(i) + b + 2M_2\mathrm{e}^{c'T}\left(\omega'(i) + \frac{b'}{c'}\right)\right] < \infty \quad \text{for all } t \in I.
\end{aligned} \tag{3.4}$$

On the other hand, by Lemma 3.1(ii) for almost every $t \in I$, we obtain

$$\begin{gathered}
\mathrm{d}\mathbb{P}^\pi_i(x_t = j) = \mathbb{E}^\pi_i\left[\int_A q(j \mid t, x_{t-}, a)\pi(\mathrm{d}a \mid e, t)\right]\mathrm{d}t, \\
\mathbb{P}^\pi_i(x_0 = j) = \delta_{ij} \quad \text{for all } i,\ j \in S.
\end{gathered} \tag{3.5}$$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Thus, using Fubini's theorem, by (3.3)–(3.5), we have

$$\begin{aligned}
&\mathbb{E}_i^\pi \int_0^T \left[\sum_{j\in S}\int_A q(j \mid s, x_s, a)\pi(\mathrm{d}a \mid e, s)\varphi(s, j)\right]\mathrm{d}s \\
&\quad = \sum_{j\in S}\int_0^T \mathbb{E}_i^\pi\left[\int_A q(j \mid s, x_{s-}, a)\pi(\mathrm{d}a \mid e, s)\right]\varphi(s, j)\,\mathrm{d}s \\
&\quad = \sum_{j\in S}\int_0^T \varphi(s, j)\,\mathrm{d}\mathbb{P}_i^\pi(x_s = j) \\
&\quad = \sum_{j\in S}\varphi(T, j)\mathbb{P}_i^\pi(x_T = j) - \varphi(0, i) - \sum_{j\in S}\int_0^T \varphi'(s, j)\mathbb{P}_i^\pi(x_s = j)\,\mathrm{d}s \\
&\quad = \mathbb{E}_i^\pi\varphi(T, x_T) - \varphi(0, i) - \mathbb{E}_i^\pi\left[\int_0^T \varphi'(s, x_s)\,\mathrm{d}s\right],
\end{aligned}$$

which implies Theorem 3.1(i).

Proof of Theorem 3.1(ii). For any $\pi_{\mathrm{t}}(\mathrm{d}a \mid \cdot) \in \Pi_{\mathrm{m}}^{\mathrm{r}}$, $s \geq t \geq 0$ and $i, j \in S$, let

$$\mathbb{P}_{ij}^\pi(t, s) := \mathbb{P}_\gamma^\pi(x_s = j \mid x_t = i).$$

Thus, by [8, Proposition C.4], for almost every $s > t$,

$$\frac{\partial \mathbb{P}_{ij}^\pi(t, s)}{\partial s} = \sum_{k\in S}\mathbb{P}_{ik}^\pi(t, s)q(j \mid s, k, \pi_s), \qquad \mathbb{P}_{ij}^\pi(t, t) = \delta_{ij}. \tag{3.6}$$

Therefore, using Fubini's theorem, by (3.3), (3.4), and (3.6), we have

$$\begin{aligned}
&\mathbb{E}_{t,i}^\pi\left[\sum_{j\in S}\int_t^T \varphi(s, j)q(j \mid s, x_s, \pi_s)\,\mathrm{d}s\right] \\
&\quad = \sum_{j\in S}\int_t^T \varphi(s, j)\sum_{k\in S}\mathbb{P}_{ik}^\pi(t, s)q(j \mid s, k, \pi_s)\,\mathrm{d}s \\
&\quad = \sum_{j\in S}\int_t^T \varphi(s, j)\,\mathrm{d}\mathbb{P}_{ij}^\pi(t, s) \\
&\quad = \sum_{j\in S}\varphi(T, j)\mathbb{P}_{ij}^\pi(t, T) - \varphi(t, i) - \sum_{j\in S}\int_t^T \varphi'(s, j)\mathbb{P}_{ij}^\pi(t, s)\,\mathrm{d}s \\
&\quad = \mathbb{E}_{t,i}^\pi\varphi(T, x_T) - \varphi(t, i) - \mathbb{E}_{t,i}^\pi\left[\int_t^T \varphi'(s, x_s)\,\mathrm{d}s\right],
\end{aligned}$$

which completes the proof.

Theorem 3.2. *Under Assumptions 3.1 and 3.2, the following assertions hold.*

(i) *If there exists* $\varphi \in C_{\omega,\omega'}^{1,0}(I \times S)$ *such that*

$$\varphi'(t, i) + r(t, i, a) + \sum_{j\in S}\varphi(t, j)q(j \mid t, i, a) \leq 0 \quad \text{for all } t \in L_\varphi^c(i),\ a \in A(t, i),$$

$$\varphi(T, i) = g(T, i), \tag{3.7}$$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

then

(ia) $V_\pi(0,i) \le \varphi(0,i)$ *for all* $\pi \in \Pi$ *and* $i \in S$;

(ib) $V_\pi(t,i) \le \varphi(t,i)$ *for all* $\pi \in \Pi^{\mathrm{r}}_{\mathrm{m}}$ *and* $(t,i) \in I \times S$.

(ii) *For any Markov policy* $f \in \Pi^{\mathrm{d}}_{\mathrm{m}}$, $V_f(\cdot,\cdot)$ *is a unique solution in* $C^{1,0}_{\omega,\omega'}(I \times S)$ *of the following equation:*

$$\varphi'(t,i) + r(t,i,f(t,i)) + \sum_{j\in S} \varphi(t,j) q(j \mid t,i,f(t,i)) = 0 \quad \text{for all } t \in L^c_\varphi(i),$$
$$\varphi(T,i) = g(T,i). \tag{3.8}$$

Proof of Theorem 3.2(i). Since $\bigcup_{i\in S} L_\varphi(i)$ is a set of Lebesgue-measure 0, by the conditions for Theorem 3.2(i) and Theorem 3.1(i), we have

$$\begin{aligned}
\mathbb{E}^\pi_i g(T,x_T) - \varphi(0,i) &= \mathbb{E}^\pi_i \varphi(T,x_T) - \varphi(0,i) \\
&= \mathbb{E}^\pi_i \left[\int_0^T (\varphi'(s,x_s) + \sum_{j\in S} \int_A \varphi(s,j) q(j \mid s,x_s,a) \pi(\mathrm{d}a \mid e,s))\, \mathrm{d}s \right] \\
&\le -\mathbb{E}^\pi_i \left[\int_0^T \int_A r(s,x_s,a) \pi(\mathrm{d}a \mid e,s)\, \mathrm{d}s \right],
\end{aligned}$$

and so

$$\mathbb{E}^\pi_i \left[\int_0^T \int_A r(s,x_s,a) \pi(\mathrm{d}a \mid e,s)\, \mathrm{d}s \right] + \mathbb{E}^\pi_i g(T,x_T) \le \varphi(0,i),$$

which implies Theorem 3.2(ia).

Similarly, by Theorem 3.1(ii) we see that Theorem 3.2(ib) also holds.

Proof of Theorem 3.2(ii). Since this proof needs similar arguments as in the proof of Theorem 4.1 below, we postpone this proof until the end of the proof of Theorem 4.1 in Section 4.

The existence of $\varphi \in C^{1,0}_{\omega,\omega'}(I \times S)$ satisfying (3.7) will be shown in Proposition 4.1 and Theorem 4.1, the proofs of which are based on some facts in Lemma 3.3 below. To state the lemma, we need some concepts. First, recall that the projection of a Borel set may not be Borel measurable but is an analytic set. Here, a subset of the Borel space X is said to be analytic (by [2, Proposition 7.41]) if it is a projection into X of a Borel subset of $X \times Y$ for some uncountable Borel space Y. Then a function $u(\cdot)$ on X is called upper semianalytic if $\{x \in X : u(x) > \delta\}$ is an analytic set for each $\delta \in (-\infty, \infty)$. It is known that each Bore measurable function is upper semianalytic; see [2, Chapter 7] for more details. Hence, $r(t,i,a)$ is upper semianalytic on $\mathbb{K}$.

Lemma 3.3. *Suppose that Assumption 3.1 holds. For any* $u \in B_\omega(I \times S)$, *define a corresponding function* $u^* : I \times S \longrightarrow (-\infty, \infty)$ *by*

$$u^*(t,i) := \sup_{a\in A(t,i)} \left\{ r(t,i,a) + \sum_{j\in S} u(t,j) q(j \mid t,i,a) \right\}.$$

Then the following assertions hold.

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

(i) *The function u^* is upper semianalytic (and hence universally measurable).*

(ii) *For every $\varepsilon > 0$, there exists a Markov policy $f \in \Pi_{\mathrm{m}}^{\mathrm{d}}$ (depending on ε) such that*

$$r(t,i,f(t,i)) + \sum_{j\in S} u(t,j)q(j \mid t,i,f(t,i)) \geq u^*(t,i) - \varepsilon \quad \textit{for all } (t,i) \in I \times S.$$

Proof. Let $Q(j \mid t,i,a) := (q(j \mid t,i,a)/q^*(i) + 1) + \delta_{ij}$ with $q^*(i)$ as in (2.2). Obviously, it is a Borel measurable stochastic kernel on S given $\mathbb{K}$. Then, by [2, Proposition 7.29], we see that $\sum_{j\in S} u(t,j)Q(j \mid t,i,a)$ is Borel measurable. Also, from

$$\sum_{j\in S} u(t,j)q(j \mid t,i,a) = \sum_{j\in S} u(t,j)Q(j \mid t,i,a)(q^*(i)+1) - u(t,i)(q^*(i)+1),$$

we can conclude that $\sum_{j\in S} u(t,j)q(j \mid t,i,a)$ is Borel measurable and, hence, $r(t,i,a) + \sum_{j\in S} u(t,j)q(j \mid t,i,a)$ is upper semianalytic. Since $u^*(t,i)$ is real-valued (by Assumption 3.1), the statements in (i) and (ii) follow from [2, Propositions 7.47 and 7.50], respectively.

4. The existence of optimal Markov policies

In this section we prove the existence of ε-optimal Markov policies and of a solution to the optimality (dynamic programming) equation (4.1) for the finite-horizon CTMDPs. The proofs are shown in two steps as follows. We first consider the case of bounded transition rates and then deal with the case of unbounded transition rates by approximations from bounded transition rates to unbounded transition rates.

The result for the case of bounded transition rates is given in the following proposition.

Proposition 4.1. *Suppose that the transition rates are bounded (i.e. $\sup_{i\in S} q^*(i) < \infty$) and Assumption 3.1 is satisfied. Then the following assertions hold.*

(i) *There exists a unique φ in $C^{1,0}_{\omega,\omega}(I \times S)$ satisfying the following optimality equation for the finite-horizon CTMDPs:*

$$\varphi'(t,i) + \sup_{a\in A(t,i)}\left[r(t,i,a) + \sum_{j\in S}\varphi(t,j)q(j \mid t,i,a)\right] = 0 \quad \textit{for all } t \in L^c_\varphi(i),$$
$$\varphi(T,i) = g(T,i). \tag{4.1}$$

(ii) *$\varphi(t,i) = V^*(t,i)$ for all $(t,i) \in I \times S$ with $\varphi(t,i)$ as in (i) above.*

(iii) *For each $\varepsilon > 0$, there exists an ε-optimal Markov policy.*

Proof of Proposition 4.1(i). For any given $\varphi \in C^{1,0}_{\omega,\omega}(I \times S)$, let $\psi(t,i) := \mathrm{e}^{\beta t}\varphi(t,i)$ for every $(t,i) \in I \times S$ with $\beta := 2L + b + c + 1$ and $L := \sup_{i\in S} q^*(i)$. Then (4.1) can be written as

$$\mathrm{e}^{-\beta t}\psi'(t,i) - \beta \mathrm{e}^{-\beta t}\psi(t,i) + \sup_{a\in A(t,i)}\left[r(t,i,a) + \mathrm{e}^{-\beta t}\sum_{j\in S}\psi(t,j)q(j \mid t,i,a)\right] = 0,$$
$$\psi(T,i) = \mathrm{e}^{\beta T} g(T,i) \quad \text{for } t \in L^c_\varphi(i),\ i \in S.$$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Since $\sup_{a\in A(s,i)}[r(s,i,a)+\mathrm{e}^{-\beta s}\sum_{j\in S}\psi(s,j)q(j\mid s,i,a)]$ is universally measurable on $I\times S$ (by Lemma 3.3(i)), the equations above are equivalent to the integral equation

$$\psi(t,i)=\mathrm{e}^{\beta t}g(T,i)+\mathrm{e}^{\beta t}\int_t^T \sup_{a\in A(s,i)}\left[r(s,i,a)+\mathrm{e}^{-\beta s}\sum_{j\in S}\psi(s,j)q(j\mid s,i,a)\right]\mathrm{d}s.$$

Define the following operator G on $B_\omega(I\times S)$. For each $\psi\in B_\omega(I\times S)$ and $(t,i)\in I\times S$,

$$G\psi(t,i):=\mathrm{e}^{\beta t}g(T,i)+\mathrm{e}^{\beta t}\int_t^T \sup_{a\in A(s,i)}\left[r(s,i,a)+\mathrm{e}^{-\beta s}\sum_{j\in S}\psi(s,j)q(j\mid s,i,a)\right]\mathrm{d}s. \tag{4.2}$$

Note that $\sup_{a\in A(s,i)}[r(s,i,a)+\mathrm{e}^{-\beta s}\sum_{j\in S}\psi(s,j)q(j\mid s,i,a)]$ is upper semianalytic (and hence universally measurable) on $I\times S$ (by Lemma 3.3) and, thus, $G\psi(t,i)$ is well defined.

On the other hand, since $L=\sup_{i\in S}q^*(i)<\infty$, from Assumption 3.1, it follows that

$$\begin{aligned}|G\psi(t,i)| &\le |\mathrm{e}^{\beta t}g(T,i)|+\mathrm{e}^{\beta t}\int_t^T\left[|r(s,i,f(s,i))|+\sum_{j\in S}|\psi(s,j)||q(j\mid s,i,f(s,i))|\right]\mathrm{d}s\\ &\le \mathrm{e}^{\beta T}[M_1+M_1T+\|\psi\|_w T(c+b+2L)]w(i)\quad\text{for all } (t,i)\in I\times S,\end{aligned} \tag{4.3}$$

which implies that $\|G\psi\|_w<\infty$. Furthermore, from (4.2) we see that $G\psi(t,i)$ (with any fixed $i\in S$) is absolutely continuous in $t\in I$, and so it is Borel measurable. Hence, $G\psi$ is in $B_\omega(I\times S)$, i.e. $G\colon B_\omega(I\times S)\longrightarrow B_\omega(I\times S)$.

For any $\psi_1,\psi_2\in B_\omega(I\times S)$, from (4.2) and $q(S\mid s,i,a)\equiv 0$, we obtain

$$\begin{aligned}&|G\psi_1(t,i)-G\psi_2(t,i)|\\ &\quad\le \mathrm{e}^{\beta t}\int_t^T \mathrm{e}^{-\beta s}\sup_{a\in A(s,i)}\sum_{j\in S}|\psi_1(s,j)-\psi_2(s,j)||q(j\mid s,i,a)|\,\mathrm{d}s\\ &\quad\le \mathrm{e}^{\beta t}\int_t^T \mathrm{e}^{-\beta s}\|\psi_1-\psi_2\|_\omega\sup_{a\in A(s,i)}\left[\sum_{j\neq i}\omega(j)q(j\mid s,i,a)+L\omega(i)\right]\mathrm{d}s\\ &\quad\le \mathrm{e}^{\beta t}\int_t^T \mathrm{e}^{-\beta s}\|\psi_1-\psi_2\|_\omega[c\omega(i)+b+2L\omega(i)]\,\mathrm{d}s\\ &\quad\le \frac{2L+b+c}{\beta}[1-\mathrm{e}^{-\beta(T-t)}]\|\psi_1-\psi_2\|_\omega\omega(i)\\ &\quad\le \frac{2L+b+c}{\beta}\|\psi_1-\psi_2\|_\omega\omega(i).\end{aligned}$$

Hence, we obtain

$$\|G\psi_1-G\psi_2\|_\omega\le\frac{2L+b+c}{\beta}\|\psi_1-\psi_2\|_\omega=\rho\|\psi_1-\psi_2\|_\omega$$

with $\rho:=(2L+b+c)/\beta=(2L+b+c)/(2L+b+c+1)<1$.

Therefore, G is a contraction operator on the Banach space $B_\omega(I\times S)$. Let $\psi^*\in B_\omega(I\times S)$ be the fixed point of G, i.e.

$$\psi^*(t,i)=\mathrm{e}^{\beta t}g(T,i)+\mathrm{e}^{\beta t}\int_t^T \sup_{a\in A(s,i)}\left[r(s,i,a)+\mathrm{e}^{-\beta s}\sum_{j\in S}\psi^*(s,j)q(j\mid s,i,a)\right]\mathrm{d}s. \tag{4.4}$$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Let $\varphi(t,i) := \mathrm{e}^{-\beta t}\psi^*(t,i)$ for all $(t,i) \in I \times S$. Then, φ is in $B_\omega(I \times S)$, and $\varphi(t,i)$ is differentiable almost everywhere and satisfies (4.1) (by (4.4)). By (4.1) and Lemma 3.3(i) we see that $\varphi'(t,i)$ is universally measurable in t (for each $i \in S$). Moreover, since $L = \sup_{i\in S} q^*(i) < \infty$, it follows from the same argument of (4.3) that $\|\varphi'\|_w^{\mathrm{es}} \le [M_1 + \|\varphi\|_w(c+b+2L)]$. Therefore, φ is in $C^{1,0}_{\omega,\omega}(I \times S)$. Thus, we complete the proof of Proposition 4.1(i).

Proof of Propositions 4.1(ii) and 4.1(iii). Since $L = \sup_{i\in S} q^*(i) < \infty$, Assumption 3.1 implies Assumption 3.2 (by taking $\omega' := \omega$). Thus, from (4.1) and Theorem 3.2(i), it follows that

$$V_\pi(0,i) \le \varphi(0,i) \quad \text{for each } \pi \in \Pi, \qquad V_\pi(t,i) \le \varphi(t,i) \quad \text{for any } \pi \in \Pi^{\mathrm{r}}_{\mathrm{m}}. \tag{4.5}$$

Moreover, since $\varphi \in B_\omega(I \times S)$, Lemma 3.3 gives the existence of $f_\varepsilon \in \Pi^{\mathrm{d}}_{\mathrm{m}}$ such that

$$\varphi'(t,i) + r(t,i,f_\varepsilon(t,i)) + \sum_{j\in S}\varphi(t,j)q(j \mid t,i,f_\varepsilon(t,i)) \ge -\frac{\varepsilon}{T} \quad \text{for all } t \in L^c_\varphi(i),$$
$$\varphi(T,i) = g(T,i),$$

which, together with Theorem 3.1(ii) and a direct calculation, leads to

$$V_{f_\varepsilon}(t,i) \ge \varphi(t,i) - \varepsilon \quad \text{for all } (t,i) \in I \times S. \tag{4.6}$$

Therefore, since ε can be arbitrary, by (4.5) and (4.6), we have

$$\sup_{\pi\in\Pi} V_\pi(0,i) = \varphi(0,i), \qquad \sup_{\pi\in\Pi^{\mathrm{r}}_{\mathrm{m}}} V_\pi(t,i) = \varphi(t,i), \qquad V_{f_\varepsilon}(t,i) \ge \varphi(t,i) - \varepsilon$$

for all $(t,i) \in I \times S$, and so Propositions 4.1(ii) and 4.1(iii) follow.

Proposition 4.1 shows the existence of $\varepsilon(>0)$-optimal Markov policies for the case of bounded transition rates. To further establish the existence of an optimal Markov policy for possibly unbounded transition rates, we need the following conditions.

Assumption 4.1. (i) *For each $(t,i) \in I \times S$, $A(t,i)$ is compact;*

(ii) *for each $t \in I, i, j \in S$, the function $q(j \mid t,i,a)$ is continuous in $a \in A(t,i)$;*

(iii) *for each $(t,i) \in I \times S$, the functions $r(t,i,a)$ and $\sum_{j\in S}\omega(j)q(j \mid t,i,a)$ are upper semicontinuous (u.s.c.) in $a \in A(t,i)$ with ω as in Assumption 3.1.*

Remark 4.1. Assumption 4.1 is the extension of [9, Conditions 6.1 and 6.2] and [21, Condition 5] for the homogeneous model to the nonhomogeneous case of $q(j \mid t,i,a)$ and $r(t,i,a)$, and it is satisfied for the examples in [8], [9], [11], [20], [21], and [24]. Assumption 4.1 is used to find the existence of the maximum points in (4.1).

Lemma 4.1. *Under Assumptions 4.1(ii) and 4.1(iii), the function $\sum_{j\in S} q(j \mid t,i,a)u(t,j)$ is u.s.c. in $a \in A(t,i)$ for every fixed $(t,i) \in I \times S$ and $u \in B_\omega(I \times S)$.*

Proof. Following the proof of [14, Lemma 8.3.7(a)] and the argument of [7, Theorem 3.3(c)], under Assumption 4.1 we see that Lemma 4.1 holds.

We next provide the main result of this paper.

Theorem 4.1. *Under Assumptions 3.1, 3.2, and 4.1, the following assertions hold.*

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

(i) *There exists a unique* φ *in* $C^{1,0}_{\omega,\omega'}(I \times S)$ *satisfying (4.1).*

(ii) $\varphi(t,i) = V^*(t,i)$ *for all* $(t,i) \in I \times S$ *with* $\varphi(t,i)$ *as in Proposition 4.1(i).*

(iii) *There exists a Markov policy* $f^* \in \Pi^{\mathrm{d}}_{\mathrm{m}}$ *such that*

$$\varphi'(t,i) + r(t,i,f^*(t,i)) + \sum_{j\in S} \varphi(t,j) q(j \mid t,i,f^*(t,i)) = 0 \quad \textit{for all } t \in L^c_\varphi(i),\ i \in S,$$

and the Markov policy f^* *is optimal.*

Proof of Theorem 4.1(i). We prove Theorem 4.1(i) by an approximation technique and Theorem 3.2(i). Since S is denumerable, without loss of generality, we define $S := \{0, 1, \dots, n, \dots\}$. For each $n \geq 1$, $j \in S$, $t \in I$, let $S_n := \{0, 1, \dots, n\}$ and

$$q_n(j \mid t,i,a) := \begin{cases} q(j \mid t,i,a) & \text{if } i \in S_n,\ a \in A(t,i), \\ 0 & \text{otherwise.} \end{cases} \tag{4.7}$$

Thus, we obtain a sequence of models $\{\mathcal{M}_n\}$ of CTMDPs as

$$\mathcal{M}_n := \{S, A, (A(t,i), (t,i) \in I \times S), r(t,i,a), q_n(j \mid t,i,a), g(t,i)\} \qquad n = 1, 2, \dots.$$

Obviously, Assumptions 3.1, 3.2, and 4.1 still hold for the data in each model $\mathcal{M}_n$. Moreover, from (2.2) and (4.7), it follows that $\sup_{i\in S} q^*_n(i) = \max\{q^*(0), \dots, q^*(n)\} < \infty$. Then for each $n \geq 1$, by Proposition 4.1, there exists $u_n \in C^{1,0}_{\omega,\omega}(I \times S)$ satisfying (4.1) for the corresponding $\mathcal{M}_n$, i.e.

$$\begin{aligned} u_n'(t,i) + \sup_{a\in A(t,i)} \left[r(t,i,a) + \sum_{j\in S} u_n(t,j) q_n(j \mid t,i,a) \right] = 0 \quad \text{for all } t \in L^c_{u_n}(i), \\ u_n(T,i) = g(T,i). \end{aligned} \tag{4.8}$$

Thus, under Assumptions 3.1 and 4.1, [13, Proposition D.5] together with Lemma 4.1 gives the existence of a Markov policy $f_n \in \Pi^{\mathrm{d}}_{\mathrm{m}}$ such that

$$\begin{aligned} u_n'(t,i) + r(t,i,f_n(t,i)) + \sum_{j\in S} u_n(t,j) q_n(j \mid t,i,f_n(t,i)) = 0 \quad \text{for all } t \in L^c_{u_n}(i), \\ u_n(T,i) = g(T,i). \end{aligned} \tag{4.9}$$

Hence, using an argument of Theorem 3.2(i) and Lemma 3.2(ii), from (4.8) and (4.9), we have

$$|u_n(t,i)| = |V_{f_n}(t,i)| \leq (T+1) M_1 \mathrm{e}^{cT} \left(1 + \frac{b}{c}\right) \omega(i) =: D\omega(i) \quad \text{for all } n \geq 1 \tag{4.10}$$

for every $(t,i) \in I \times S$, where $D := (T+1) M_1 \mathrm{e}^{cT} (1 + b/c)$.

Next, we prove that $\{u_n,\ n \geq 1\}$ is equicontinuous on $I \times S$. To do so, let $H_n(s,i) := \sup_{a\in A(s,i)} [r(s,i,a) + \sum_{j\in S} u_n(s,j) q_n(j \mid s,i,a)]$ for every $(s,i) \in I \times S$. Then, from

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Assumptions 3.1, 3.2, and (4.10), we have

$$\begin{aligned}|H_n(s,i)| &\le \sup_{a\in A(s,i)}\left[|r(s,i,a)| + \sum_{j\in S}|u_n(s,j)||q_n(j\mid s,i,a)|\right]\\ &\le \sup_{a\in A(s,i)} e\left[M_1\omega(i) + D\sum_{j\in S}\omega(j)|q(j\mid s,i,a)|\right]\\ &= \sup_{a\in A(s,i)}\left[M_1\omega(i) + D\sum_{j\in S}\omega(j)q(j\mid s,i,a) - 2Dq(i\mid s,i,a)\omega(i)\right]\\ &\le M_1\omega(i) + D[c\omega(i) + b + 2M_2\omega'(i)]\\ &=: L(i) \quad \text{for all } (s,i)\in I\times S. \end{aligned} \tag{4.11}$$

On the other hand, note that (4.8) is equivalent to the following integral equation:

$$\begin{aligned}u_n(t,i) &= g(T,i) + \int_t^T \sup_{a\in A(s,i)}\left[r(s,i,a) + \sum_{j\in S}u_n(s,j)q_n(j\mid s,i,a)\right]\mathrm{d}s\\ &= g(T,i) + \int_t^T H_n(s,i)\,\mathrm{d}s \quad \text{for all } (t,i)\in I\times S. \end{aligned} \tag{4.12}$$

Thus, given any $(t_0,i_0)\in I\times S$ and $\varepsilon>0$, take $\delta := \min\{\varepsilon/L(i_0), \frac{1}{2}\}$. For every (t,i) in the open set $\{(t,i)\in I\times S\colon |t-t_0|<\delta, |i-i_0|<\delta\}$, we have $i=i_0$, and so (by (4.12))

$$\begin{aligned}|u_n(t,i) - u_n(t_0,i_0)| &= |u_n(t,i_0) - u_n(t_0,i_0)|\\ &= \left|\int_t^T H_n(s,i_0)\,\mathrm{d}s - \int_{t_0}^T H_n(s,i_0)\,\mathrm{d}s\right|\\ &= \left|\int_t^{t_0} H_n(s,i_0)\,\mathrm{d}s\right| \le L(i_0)|t-t_0| < \varepsilon \quad \text{for all } n\ge 1. \end{aligned}$$

Hence, $\{u_n,\ n\ge 1\}$ is equicontinuous at (t_0,i_0), which, together with the arbitrariness of $(t_0,i_0)\in I\times S$, yields that $\{u_n,\ n\ge 1\}$ is equicontinuous on $I\times S$. Thus, the Ascoli theorem (see, e.g. [13, p. 96]) gives the existence of a subsequence $\{u_{n_k}, k\ge 1\}$ of $\{u_n, n\ge 1\}$ and a continuous function φ on $I\times S$ such that

$$\lim_{k\to\infty} u_{n_k}(t,i) = \varphi(t,i), \qquad |\varphi(t,i)| \le D\omega(i) \quad \text{for all } (t,i)\in I\times S. \tag{4.13}$$

Let $H(s,i) := \sup_{a\in A(s,i)}[r(s,i,a) + \sum_{j\in S}\varphi(s,j)q(j\mid s,i,a)]$ for all $(s,i)\in I\times S$. We next show that $\lim_{k\to\infty} H_{n_k}(s,i) = H(s,i)$ for each $(s,i)\in I\times S$.

Indeed, for any fixed $(s,i)\in I\times S$, since $q_{n_k}(j\mid s,i,a)\to q(j\mid s,i,a)$ for all $j\in S$ and $a\in A(s,i)$ as $k\to\infty$ (by (4.7)), by [14, Lemma 8.3.7] and (4.10), we have

$$\begin{aligned}\liminf_{k\to\infty} H_{n_k}(s,i) &\ge \liminf_{k\to\infty}\left[r(s,i,a) + \sum_{j\in S}u_{n_k}(s,j)q_{n_k}(j\mid s,i,a)\right]\\ &\ge r(s,i,a) + \sum_{j\in S}\varphi(s,j)q(j\mid s,i,a) \quad \text{for all } a\in A(s,i). \end{aligned}$$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Hence,

$$\liminf_{k\to\infty} H_{n_k}(s,i) \ge \sup_{a\in A(s,i)}\Big[r(s,i,a) + \sum_{j\in S}\varphi(s,j)q(j \mid s,i,a)\Big]. \tag{4.14}$$

On the other hand, note that $\limsup_{k\to\infty} H_{n_k}(s,i) = \lim_{m\to\infty} H_{n_{k_m}}(s,i)$ for some subsequence $\{n_{k_m},\ m \ge 1\}$ of $\{n_k,\ k \ge 1\}$. For each $m \ge 1$, under Assumption 4.1, the measurable selection theorem (see, e.g. [13, Proposition D.5]) together with Lemma 4.1 ensures the existence of $f_{n_{k_m}} \in \Pi_{\mathrm{m}}^{\mathrm{d}}$ such that

$$\begin{aligned} H_{n_{k_m}}(s,i) &= \sup_{a\in A(s,i)}\Big[r(s,i,a) + \sum_{j\in S} u_{n_{k_m}}(s,j)q_{n_{k_m}}(j \mid s,i,a)\Big] \\ &= r(s,i,f_{n_{k_m}}(s,i)) + \sum_{j\in S} u_{n_{k_m}}(s,j)q_{n_{k_m}}(j \mid s,i,f_{n_{k_m}}(s,i)). \end{aligned} \tag{4.15}$$

Since $f_{n_{k_m}}(s,i) \in A(s,i)$ for all $m \ge 1$ and $A(s,i)$ is compact, there exists a subsequence $\{f_{n_{k_{m_l}}}(s,i),\ l \ge 1\}$ of $\{f_{n_{k_m}}(s,i), m \ge 1\}$ and $a(s,i) \in A(s,i)$ (depending on (s,i)) such that $f_{n_{k_{m_l}}}(s,i) \to a(s,i)$ as $l \to \infty$ and $\lim_{m\to\infty} H_{n_{k_m}}(s,i) = \lim_{l\to\infty} H_{n_{k_{m_l}}}(s,i)$. Thus, using Assumption 4.1, by [14, Lemma 8.3.7] and (4.15), we have

$$\begin{aligned} \limsup_{k\to\infty} H_{n_k}(s,i) &= \lim_{l\to\infty} H_{n_{k_{m_l}}}(s,i) \\ &= \lim_{l\to\infty}\Big[r(s,i,f_{n_{k_{m_l}}}(s,i)) + \sum_{j\in S} u_{n_{k_{m_l}}}(s,j)q_{n_{k_{m_l}}}(j \mid s,i,f_{n_{k_{m_l}}}(s,i))\Big] \\ &= r(s,i,a(s,i)) + \sum_{j\in S}\varphi(s,j)q(j \mid s,i,a(s,i)) \\ &\le \sup_{a\in A(s,i)}\Big[r(s,i,a) + \sum_{j\in S}\varphi(s,j)q(j \mid s,i,a)\Big], \end{aligned}$$

which, together with (4.14), implies that $\lim_{k\to\infty} H_{n_k}(s,i) = H(s,i)$ and, thus, from (4.11)–(4.13), it follows that

$$\begin{aligned} \varphi(t,i) &= g(T,i) \\ &\quad + \int_t^T \sup_{a\in A(s,i)}\Big[r(s,i,a) + \sum_{j\in S}\varphi(s,j)q(j \mid s,i,a)\Big]\mathrm{d}s \quad \text{for all } (t,i) \in I \times S. \end{aligned} \tag{4.16}$$

This implies that $\varphi(t,i)$ is differentiable almost everywhere in $t \in I$ and satisfies (4.1). By (4.1) and Lemma 3.3(i) we see that $\varphi'(t,i)$ is universally measurable in t (for each $i \in S$). To show that $\varphi \in C^{1,0}_{\omega,\omega'}(I \times S)$, since $\varphi \in B_\omega(I \times S)$ (just proved in (4.13)), the rest of the proof verifies that φ' is essentially $(w + w')$-bounded on $I \times S$. Indeed, as the arguments in (4.11), from (4.1), we have

$$\begin{aligned} |\varphi'(t,i)| &\le M_1\omega(i) + \|\varphi\|_\omega[c\omega(i) + b + 2M_2\omega'(i)] \\ &\le [M_1 + \|\varphi\|_\omega(c + b + 2M_2)](\omega(i) + \omega'(i)) \quad \text{for all } t \in L^c_\varphi(i)\, i \in S. \end{aligned}$$

Hence, we have $\|\varphi'\|^{\mathrm{es}}_{w+w'} \le M_1 + \|\varphi\|_\omega(c + b + 2M_2) < \infty$. Therefore, φ is in $C^{1,0}_{\omega,\omega'}(I \times S)$ and, thus, we complete the proof of Theorem 4.1(i).

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

Proof of Theorems 4.1(ii) and 4.1(iii). From (4.1), it follows that

$$\varphi'(t,i)+r(t,i,a)+\sum_{j\in S}\varphi(t,j)q(j\mid t,i,a)\le 0 \quad \text{for all } t\in L^c_\varphi(i),\ a\in A(t,i),$$
$$\varphi(T,i)=g(T,i). \tag{4.17}$$

Using Theorem 3.2(i), by (4.17), we have for each $t\in I, i\in S$,

$$V_\pi(0,i)\le\varphi(0,i) \quad \text{for each } \pi\in\Pi, \qquad V_\pi(t,i)\le\varphi(t,i) \quad \text{for any } \pi\in\Pi^{\mathrm{r}}_{\mathrm{m}}. \tag{4.18}$$

Moreover, since $\varphi\in B_\omega(I\times S)$, the measurable selection theorem (see, e.g. [13, Proposition D.5]) together with Lemma 4.1 gives the existence of $f^*\in\Pi^{\mathrm{d}}_{\mathrm{m}}$ such that

$$\varphi'(t,i)+r(t,i,f^*(t,i))+\sum_{j\in S}\varphi(t,j)q(j\mid t,i,f^*(t,i))=0 \quad \text{for all } t\in L_\varphi(i),$$
$$\varphi(T,i)=g(T,i),$$

which, together with an argument of Theorem 3.2(i), gives $V_{f^*}(t,i)=\varphi(t,i)$ for all $(t,i)\in I\times S$. Therefore, by (4.18), we have

$$\sup_{\pi\in\Pi}V_\pi(0,i)=V_{f^*}(0,i)=\varphi(0,i), \qquad \sup_{\pi\in\Pi^{\mathrm{r}}_{\mathrm{m}}}V_\pi(t,i)=V_{f^*}(t,i)=\varphi(t,i)$$

for all $(t,i)\in I\times S$, and so Theorems 4.1(ii) and 4.1(iii) follow.

Proof of Theorem 3.2(ii). For any given Markov policy f, we first show the existence of a $\varphi\in C^{1,0}_{\omega,\omega'}(I\times S)$ satisfying (3.8) in the following two steps.

Step 1. (On the assumption that $\sup_{i\in S}q^*(i)<\infty$.) We modify the operator G in (4.2) as the following G^f:

$$G^f\psi(t,i):=\mathrm{e}^{\beta t}g(T,i)+\mathrm{e}^{\beta t}\int_t^T\left[r(s,i,f(s,i))+\mathrm{e}^{-\beta s}\sum_{j\in S}\psi(s,j)q(j\mid s,i,f(s,i))\right]\mathrm{d}s$$

for all $(t,i)\in I\times S$. Then a similar argument as in the proof of Proposition 4.1 gives the existence of $\varphi\in C^{1,0}_{\omega,\omega}(I\times S)$ satisfying (3.8).

Step 2. (The approximation technique.) For each $n\ge 1$, $j\in S$, $t\in I$, from (4.7), it follows that

$$q_n(j\mid t,i,f(t,i))=\begin{cases}q(j\mid t,i,f(t,i)) & \text{if } 0\le i\le n,\\ 0 & \text{otherwise.}\end{cases}$$

Thus, for each $n\ge 1$, by step 1 above, there exists $v_n\in C^{1,0}_{\omega,\omega}(I\times S)$ (depending on f) satisfying (3.8) with $q(j\mid t,i,f(t,i):=q_n(j\mid t,i,f(t,i))$. As in the arguments for Theorem 4.1(i), $\{v_n,\ n\ge 1\}$ is equicontinuous on $I\times S$. Thus, there exists a subsequence $\{v_{n_k},k\ge 1\}$ of $\{v_n,n\ge 1\}$ and a continuous function $\hat\varphi$ on $I\times S$ such that $\lim_{k\to\infty}v_{n_k}(t,i)=\hat\varphi(t,i)$ for all $(t,i)\in I\times S$. Furthermore, as in the proof of (4.16), we have

$$\hat\varphi(t,i)=g(T,i)+\int_t^T\left[r(s,i,f(s,i))+\sum_{j\in S}\hat\varphi(s,j)q(j\mid s,i,f(s,i))\right]\mathrm{d}s \quad \text{for all } (t,i)\in I\times S,$$

which implies that $\hat\varphi$ is in $C^{1,0}_{\omega,\omega'}(I\times S)$ and satisfies (3.8).

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

We now prove the uniqueness of a solution to (3.8). Suppose that a function $\varphi \in C^{1,0}_{\omega,\omega'}(I \times S)$ satisfies (3.8). For each $(t, i) \in I \times S$, since $\bigcup_{i\in S} L_\varphi(i)$ is a set of Lebesgue-measure 0, by Theorem 3.1(ii), we have

$$\begin{aligned}\mathbb{E}^f_{t,i} g(T, x_T) - \varphi(t, i) &= \mathbb{E}^f_{t,i}\varphi(T, x_T) - \varphi(t, i)\\ &= \mathbb{E}^f_{t,i}\left[\int_t^T \left(\varphi'(s, x_s) + \sum_{j\in S} \varphi(s, j) q(j \mid s, x_s, f(s, x_s))\right) \mathrm{d}s\right]\\ &= -\mathbb{E}^f_{t,i}\left[\int_t^T r(s, x_s, f(s, x_s))\, \mathrm{d}s\right],\end{aligned}$$

and so

$$\varphi(t, i) = \mathbb{E}^f_{t,i}\left[\int_t^T r(s, x_s, f(s, x_s))\, \mathrm{d}s\right] + \mathbb{E}^f_{t,i} g(T, x_T) = V_f(t, i),$$

which implies the uniqueness. Thus, the proof of Theorem 3.2(ii) is completed.

Remark 4.2. Theorem 4.1 and Proposition 4.1 establish the existence of an optimal Markov policy and an ε-optimal Markov policy, respectively. Moreover, they allow the value function $V^*(t, i)$ to be *unbounded*; see Proposition 5.1 and Remark 5.1 below for more details. This shows that the required boundedness of the value function in [27, Theorems 5.1 and 5.2] with *bounded transition rates* can be done away with and, thus, the corresponding unsolved problems in [27, Theorems 5.1 and 5.2, p. 234] have been solved for the case of finite-horizon CTMDPs.

5. An example

Recall that Assumptions 3.1, 3.2, and 4.1 above are generalizations of those in [8], [11], [20], [21], and [24]. Hence, all examples in these references satisfy these assumptions. To further illustrate the main results here, we provide an example.

Example 5.1. (*A controlled birth–death system.*) Consider a birth and death system in which the state variable denotes the population size at time t. The 'natural' birth and death rates at time $t \geq 0$ are denoted by $\lambda(t)$ and $\mu(t)$, respectively. Suppose that there are additional birth and death parameters denoted by a_1 and a_2, respectively, which are assumed to be controlled by a decision maker. When the state of the system is $i \geq 0$, the decision maker takes an action (a_1, a_2) from a given set $A_1(i) \times A_2(i)$, which may increase (i.e. $a_1, a_2 \geq 0$) or decrease (i.e. $a_1, a_2 \leq 0$) the birth (death) rate. The action results in a reward $r(t, i, a_1, a_2)$, and also affects the birth–death rates given by (5.1) and (5.2) below.

The model for this birth–death system is as follows:

- the state space is $S = \{0, 1, \ldots, i, \ldots\}$;
- the action A equals to $\bigcup_{i\in S} A(t, i)$ with $A(t, i) := A_1(i) \times A_2(i)$ for all $t \geq 0$ and $i \in S$;
- the transition rates $q(j \mid t, i, a)$ (with $a := (a_1, a_2)$) are given by (5.1) and (5.2) below, and the reward is $r(t, i, a) := r(t, i, a_1, a_2)$ for every $t \geq 0, i \in S, a = (a_1, a_2) \in A(t, i)$.

When $i = 0$, there is no population in the system, and so it is natural to set $A_2(0) := \{0\}$. Thus, for each $t \geq 0$, we have

$$q(1 \mid t, 0, a) = -q(0 \mid t, 0, a) := \lambda(t) + a_1 \quad \text{for all } a := (a_1, a_2) \in A_1(0) \times A_2(0), \quad (5.1)$$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

where a_1 is explained as an immigration parameter. For each $t \geq 0$, $i \geq 1$, and $a = (a_1, a_2) \in A_1(i) \times A_2(i)$,

$$q(j \mid t, i, a) := \begin{cases} \lambda(t)i + a_1 & \text{if } j = i+1, \\ -[\lambda(t) + \mu(t)]i - a_1 - a_2 & \text{if } j = i, \\ \mu(t)i + a_2 & \text{if } j = i-1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

The aim here is to find conditions under which there exists an optimal Markov policy for the CTMDPs with any given horizon $T > 0$ and the terminal reward g. To do so, we consider the following hypotheses.

(C1) $\lambda(t)$ and $\mu(t)$ are continuous, nonnegative, and bounded in $t \geq 0$. (Hence, the constants $\lambda_1 := \inf_{t\geq 0} \lambda(t)$, $\lambda_2 := \sup_{t\geq 0} \lambda(t)$, $\mu_1 := \inf_{t\geq 0} \mu(t)$, and $\mu_2 := \sup_{t\geq 0} \mu(t)$ are all nonnegative and finite.)

(C2) $A_1(i)$ is a closed subset of $[-\lambda_1 i, (k + \lambda_2)(1 + i)]$ for each $i \geq 0$ with some integer $k \geq 1$; and $A_2(i)$ is a closed subset of $[-\mu_1 i, (2 + \mu_2)(1 + i)]$ for each $i \geq 1$.

(C3) For each $(t, i) \in I \times S$, the function $r(t, i, a)$ is u.s.c. in $a \in A(t, i)$ and there exists a constant $M > 0$ such that $|g(T, i)| \leq M(i^n + 1)$ and $|r(t, i, a)| \leq M(i^n + 1)$ for all $t \in I$, $i \in S$ and $a \in A(t, i)$, where $n \geq 1$ is some integer.

Under these conditions, we obtain the following proposition.

Proposition 5.1. *Under (C1), (C2), and (C3), the following assertions hold (for Example 5.1).*

(i) *The controlled birth–death system satisfies Assumptions 3.1, 3.2, and 4.1. Therefore (by Theorem 4.1), there exists an optimal Markov policy.*

(ii) *(Special case 1.) Suppose that, in addition,* $\lambda(t) = \mu(t) = 0$ *for all* $t \geq 0$, $A_1(i) = [0, i]$, $A_2(i) = [0, 2i]$ *for all* $i \in S$; *the reward functions* $r(t, i, a_1, a_2)$ *and* $g(t, i)$ *are given by* $r(t, i, a_1, a_2) = -2i + (T + 3 - 3e^{t-T/2})a_1 + (\frac{3}{2}e^{t-T/2} - \frac{3}{2} - T)a_2$ *for* $t \in [0, T/2)$ *and* $r(t, i, a_1, a_2) = -2i + (5T/2 - 3t)a_1 + (t - 3T/2)a_2$ *for* $t \in [T/2, T]$, *where* $a_1 \in [0, i]$, $a_2 \in [0, 2i]$, *and* $g(T, i) = 0$ *for all* $i \geq 0$. *Then, for every* $i \geq 0$, *the value function* $V^*(t, i)$ *and an optimal Markov policy* $f^*(t, i)$ *are given as*

$$V^*(t, i) = \begin{cases} -i(2 + T - 2e^{t-T/2}), & t \in [0, T/2), \\ -2i(T - t), & t \in [T/2, T], \end{cases} \tag{5.3}$$

$$f^*(t, i) = \begin{cases} (i, 2i), & t \in [0, T/2), \\ (0, 0), & t \in [T/2, T]. \end{cases} \tag{5.4}$$

(iii) *(Special case 2.) Suppose that, in addition,* $\lambda(t) = \mu(t) = 0$ *for all* $t \geq 0$, $A_1(0) = [0, k]$, $A_2(0) = \{0\}$, *and* $A_1(i) = [0, i]$, $A_2(i) = [0, 2i]$ *for all* $i \geq 1$; *the reward functions*

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

$r(t,i,a_1,a_2)$ and $g(T,i)$ are defined by

$$r(t,i,a_1,a_2)=\begin{cases}(-2\mathrm{e}^{t-T/2}-kt)a_1, & i=0,\ t\in[0,\ T/2),\\ (T-2-2t-kt)a_1, & i=0,\ t\in[T/2,T],\\ -4+(2+T)a_1+(3+kt)a_2, & i=1,\ t\in[0,T/2),\\ -2, & i=1,\ t\in[T/2,T],\\ -2i-Ta_1+\left(\dfrac{3}{2}\mathrm{e}^{t-T2}-\dfrac{3}{2}-T\right)(a_2-2a_1), & i>1,\ t\in[0,T/2),\\ -2i+\left(\dfrac{5T}{2}-3t\right)a_1+\left(t-\dfrac{3T}{2}\right)a_2, & i>1,\ t\in[T/2,T],\end{cases}$$

and $g(T,i)=0$ for all $i\geq 0$. Then the value function $V^*(t,i)$ and an optimal Markov policy $f^*(t,i)$ are given as

$$V^*(t,i)=\begin{cases}-kt-3-T, & i=0,\ t\in[0,T),\\ -i(2+T-2\mathrm{e}^{t-T/2}), & i\geq 1,\ t\in[0,T/2),\\ -2i(T-t), & i\geq 1,\ t\in[T/2,T],\end{cases}\tag{5.5}$$

and

$$f^*(t,i)=\begin{cases}(k,0), & i=0,\ t\in[0,T),\\ (i,2i), & i\geq 1,\ t\in[0,T/2),\\ (0,0), & i\geq 1,\ t\in[T/2,T),\end{cases}\tag{5.6}$$

respectively.

Proof of Proposition 5.1(i). We shall first verify Assumption 3.1. Let $\omega(i):=i^n+1$ for each $i\in S$, and $S_m:=\{0,1,\dots,m\}$ for all $m\geq 1$, where n is the same as in (C3). It is obvious that $S_m\uparrow S$, $\sup_{i\in S_m}q^*(i)<\infty$ and $\lim_{m\to\infty}\inf_{j\notin S_m}\omega(j)=\lim_{m\to\infty}[(m+1)^n+1]=+\infty$. Moreover, for each $i\geq 1$ and $a=(a_1,a_2)\in A(t,i)$, using (C1) and (C2), by (5.2), we have

$$\begin{aligned}\sum_{j\in S}q(j\mid t,i,a)\omega(j)&=[a_2+\mu(t)i](i-1)^n-[a_1+a_2+\mu(t)i+\lambda(t)i]i^n\\ &\quad+(a_1+\lambda(t)i)(i+1)^n\\ &\leq 2^n(\lambda(t)i+a_1)i^{n-1}\\ &\leq 2^n[\lambda_2+2(k+\lambda_2)]\omega(i)\quad\text{for all } t\geq 0.\end{aligned}\tag{5.7}$$

For $i=0$, we have

$$\sum_{j\in S}q(j\mid t,0,a)\omega(j)=\lambda(t)+a_1\leq 2^n[\lambda_2+2(k+\lambda_2)]\omega(0).\tag{5.8}$$

From (5.7) and (5.8) we conclude that Assumption 3.1 holds under (C1)–(C3).

By (C1)–(C3) and (5.2), Assumption 4.1 is obviously satisfied. Moreover, take $\omega'(i):=i^{n+1}+1$ for all $i\in S$. Then, as in the proofs of (5.7) and (5.8), we can derive that Assumption 3.2 also holds. Hence, Assumptions 3.1, 3.2, and 4.1 are verified.

Proof of Proposition 5.1(ii). Under the conditions in Proposition 5.1(ii), by modifying $\varphi'(t,i)$ on $L_\varphi(i)$ in obvious ways, (4.1) can be expressed as

$$\varphi'(t,0)=0,\qquad \varphi(T,0)=0,\tag{5.9}$$

https://doi.org/10.1239/aap/1449859800 Published online by Cambridge University Press

and

$$\sup_{a\in A(t,i)}[r(t,i,a)+a_2\varphi(t,i-1)-(a_1+a_2)\varphi(t,i)+a_1\varphi(t,i+1)]=-\varphi'(t,i),$$
$$\varphi(T,i)=0 \quad \text{for every } i\geq 1,\ t\in[0,T], \tag{5.10}$$

where $a=(a_1,a_2)$ and $A(t,i)=[0,i]\times[0,2i]$.

By solving (5.9) and (5.10), we obtain the value function $V^*(t,i)$ as in (5.3). Furthermore, from (5.10) we derive an optimal policy $f^*(t,i)$ as in (5.4).

Proof of Proposition 5.1(iii). Under the conditions in Proposition 5.1(iii), by modifying $\varphi'(t,i)$ on $L_\varphi(i)$ in obvious ways, (4.1) can be expressed as

$$\varphi'(t,0)+\sup_{a\in[0,k]\times\{0\}}[r(t,0,a)-a_1\varphi(t,0)+a_1\varphi(t,1)]=0, \qquad \varphi(T,0)=0, \tag{5.11}$$

and

$$\sup_{a\in A(t,i)}[r(t,i,a)+a_2\varphi(t,i-1)-(a_1+a_2)\varphi(t,i)+a_1\varphi(t,i+1)]=-\varphi'(t,i),$$
$$\varphi(T,i)=0 \quad \text{for every } i\geq 1,\ t\in[0,T], \tag{5.12}$$

where $a=(a_1,a_2)$ and $A(t,i)=[0,i]\times[0,2i]$.

By solving (5.11) and (5.12), we obtain the value function $V^*(t,i)$ as in (5.5). Furthermore, using the value function and (5.11) and (5.12), we obtain an optimal policy $f^*(t,i)$ as in (5.6).

Remark 5.1. Although the value function $V^*(t,i)$ in Example 5.1 is finite (by Theorem 4.1), from (5.3) (or (5.5)) we see that it can be *unbounded* since $\inf_{i\in S}V^*(t,i)=-\infty$ for each $t\in[0,T)$. This implies that the required boundedness of the value function in [27, Theorems 5.1 and 5.2] can be done away with and, thus, we have obtained solutions to the unsolved problems for the finite-horizon CTMDPs in [27, Theorems 5.1 and 5.2, p. 234].

Acknowledgements

The authors thank the AE and the anonymous referee for their numerous valuable comments and suggestions which have improved this paper. This research was partially supported by the National Natural Science Foundation of China and the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme.

References

[1] Baüerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance.* Springer, Heidelberg.
[2] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control. The Discrete Time Case.* Academic Press, New York.
[3] Feinberg, E. A. (2004). Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Operat. Res.* **29,** 492–524.
[4] 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*. Birkhäuser, New York, pp. 77–97.
[5] 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.
[6] Ghosh, M. K. and Saha, S. (2012). Continuous-time controlled jump Markov processes on the finite horizon. In *Optimization, Control, and Applications of Stochastic Systems*. Birkhäuser, New York, pp. 99–109.
[7] Guo, X. (2007). Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces. *Math. Operat. Res.* **32,** 73–87.

[8] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes.* Springer, Berlin.
[9] Guo, X. and Piunovskiy, A. (2011). Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. *Math. Operat. Res.* **36,** 105–132.
[10] Guo, X. and Ye, L. (2010). New discount and average optimality conditions for continuous-time Markov decision processes. *Adv. Appl. Prob.* **42,** 953–985.
[11] Guo, X., Hernández-Lerma, O. and Prieto-Rumeau, T. (2006). A survey of recent results on continuous-time Markov decision processes. *Top* **14,** 177–261.
[12] Guo, X., Huang, Y. and Song, X. (2012). Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent policies. *SIAM J. Control Optimization* **50,** 23–47.
[13] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes. Basic Optimality Criteria.* Springer, New York.
[14] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes.* Springer, New York.
[15] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon–Nikodým derivatives, representation of martingales. *Z. Wahrscheinlichkeitsth.* **31,** 235–253.
[16] Kakumanu, P. (1971). Continuously discounted Markov decision model with countable state and action space. *Ann. Math. Statist.* **42,** 919–926.
[17] Kakumanu, P. (1975). Continuous time Markovian decision processes average return criterion. *J. Math. Anal. Appl.* **52,** 173–188.
[18] Kitaev, M. Y. and Rykov, V. V. (1995). *Controlled Queueing Systems.* CRC Press, Boca Raton, FL.
[19] Miller, B. L. (1968). Finite state continuous time Markov decision processes with a finite planning horizon. *SIAM. J. Control* **6,** 266–280.
[20] Piunovskiy, A. and Zhang, Y. (2011). Accuracy of fluid approximations to controlled birth-and-death processes: absorbing case. *Math. Meth. Operat. Res.* **73,** 159–187.
[21] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optimization* **49,** 2032–2061.
[22] Pliska, S. R. (1975). Controlled jump processes. *Stoch. Process. Appl.* **3,** 259–282.
[23] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). Discounted continuous-time controlled Markov chains: convergence of control models. *J. Appl. Prob.* **49,** 1072–1090.
[24] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games.* Imperial College Press, London.
[25] Prieto-Rumeau, T. and Lorenzo, J. M. (2010). Approximating ergodic average reward continuous-time controlled Markov chains. *IEEE Trans. Automatic Control* **55,** 201–207.
[26] Ye, L. and Guo, X. (2012). Continuous-time Markov decision processes with state-dependent discount factors. *Acta Appl. Math.* **121,** 5–27.
[27] Yushkevich, A. A. (1978). Controlled Markov models with countable state space and continuous time. *Theory Prob. Appl.* **22,** 215–235.