J. Appl. Prob. **51,** 954–970 (2014)
Printed in England

AVERAGE OPTIMALITY FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES UNDER WEAK CONTINUITY CONDITIONS

YI ZHANG,* *University of Liverpool*

Abstract

This paper considers the average optimality for a continuous-time Markov decision process in Borel state and action spaces, and with an arbitrarily unbounded nonnegative cost rate. The existence of a deterministic stationary optimal policy is proved under the conditions that allow the following; the controlled process can be explosive, the transition rates are weakly continuous, and the multifunction defining the admissible action spaces can be neither compact-valued nor upper semicontinuous.

Keywords: Continuous-time Markov decision process; average optimality; weak continuity

2010 Mathematics Subject Classification: Primary 90C40
Secondary 60J25

1. Introduction

In this paper we establish the existence of a deterministic stationary average optimal policy for a possibly explosive CTMDP (continuous-time Markov decision process) in Borel state and action spaces under the weak continuity condition.

The average criterion for CTMDPs has been studied by many authors; for the recent developments; see [14], [16], [20], and [32] for the case of a countable state space, and [17], [18], [28], and [34] for the case of a possibly uncountable state space. Considering a nonnegative cost rate, as in this paper, the standard approach of proving the existence of a deterministic stationary optimal policy for an average CTMDP is through the optimality inequality [16], [18]. If additional, but less verifiable, conditions are imposed, we can establish the optimality equation [14], [34]. In general, it is known [16] that the optimality equation may not have a solution even if the optimality inequality can be solved; see also [3].

In this paper, for the CTMDP with Borel state and action spaces, and a nonnegative cost rate, we also follow the optimality inequality approach, however, under different conditions from the present literature on CTMDPs with the average criterion. We will explain that our conditions are rather general, in which the contribution of this paper also lies.

Firstly, all the aforementioned works on CTMDPs [14], [16], [17], [18], [20], [28], [32], [34] assume the underlying process to be nonexplosive; and most of them achieve this by assuming the existence of a Lyapunov function bounding the growth of the transition rates. In this paper we remove this condition, and allow the transition rates to be essentially arbitrarily unbounded, and the controlled process to be possibly explosive. The development of the theory covering such CTMDPs was once regarded as quite challenging in the survey by Guo *et al.* [20]; for the discounted criteria, we refer the reader to [6] and [31].

Received 19 July 2013; revision received 18 November 2013.
* Postal address: Department of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, UK.
Email address: yi.zhang@liv.ac.uk

Secondly, we assume the weak continuity on the underlying signed kernel defining the transition rates, while all the previous literature on average CTMDPs in Borel spaces was based on the strong continuity condition, except for [21], which established the existence of a randomized stationary optimal policy for the constrained CTMDPs. It is relevant to point out that recently the developments of the theory of average DTMDPs (discrete-time Markov decision processes) and SMDPs (semi-Markov decision processes) with weakly continuous (also called Feller) transition probabilities have received much attention from the research community [5], [7], [8], [25], [26], [24]. In a nutshell, as compared to the strongly continuous case, the proofs with weakly continuous transition rates are more technical, and the construction of the solution to the optimality inequality would involve the notion of the generalized lower limit and the generalized Fatou's lemma. Moreover, based on a neat generalization of the Berge theorem [9], which is partially summarized in Lemma 3 of this paper, and as in [8] for the average DTMDP, we allow the multifunction defining the admissible action spaces to be neither compact-valued nor upper semicontinuous.

If the state space is countable, then the concepts of weak and strong continuity coincide. However, in general, meaningful applications of Markov control problems to, for example, inventory management, have been noted, where the weak continuity condition can be satisfied, while the strong continuity condition is not; see the examples in [24, Section 6].

Since the solution to the optimality inequality is constructed following the vanishing discount factor approach, some of the results about discounted CTMDPs are incidentally extended in this paper as well.

Out of the current literature on CTMDPs, this paper is most closely related to [18], which is an extension of [16], and also derives the average optimality inequality for a CTMDP. Nevertheless, it assumes the existence of a Lyapunov function, and considers strongly continuous transition rates. A more detailed comparison of our conditions with those of [18] is presented after Condition 4.

Finally, since we allow the transition rates to be essentially arbitrarily unbounded and not separated from zero, the standard technique transforming the concerned average CTMDP to an equivalent DTMDP [33] remains to be formally justified and is not directly applicable to our setup.

The rest of this paper is organized as follows. In Section 2 we describe the concerned CTMDP problem. In Section 3 we present the main result. The proof of the main result is postponed until Section 4, with some auxiliary statements being presented therein. We conclude this paper with Section 5. To improve the readability, the proofs of the auxiliary results and some definitions, together with known lemmas, are collected in Appendix A.

2. Optimal control problem statement

In what follows, $\mathbf{1}$ 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. Below, unless stated otherwise, the term of measurability is always understood in the Borel sense, and a function can take values in $[-\infty, \infty]$. The convention of $\infty - \infty := \infty$ is in use.

The primitives of a CTMDP are the elements $\{S, A, (A(x) \subseteq A,\ x \in S), q(\cdot \mid x, a)\}$, where S is a nonempty Borel state space, i.e. a measurable subset of some complete separable metric space, A is a nonempty Borel action space, and the multifunction $A(\cdot)\colon x \mapsto A(x) \subseteq A$ specifies the admissible action spaces, for which we assume that $A(x) \in \mathcal{B}(A)$ for each $x \in S$, and its graph $K := \{(x, a)\colon x \in S,\ a \in A(x)\}$ belongs to $\mathcal{B}(S \times A)$ and contains the graph of at least one measurable mapping from S to A. This assumption guarantees the existence of the

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

deterministic stationary policies defined below. The transition rates are given by $q(\cdot \mid x, a)$, a signed kernel on $\mathcal{B}(S)$ given that $(x, a) \in K$ such that $q(\Gamma_S \setminus \{x\} \mid x, a) \geq 0$ for all $\Gamma_S \in \mathcal{B}(S)$. Throughout this paper we assume that $q(\cdot \mid x, a)$ is conservative and stable, i.e. $q(S \mid x, a) = 0$ and $\bar{q}_x = \sup_{a \in A(x)} q_x(a) < \infty$, where $q_x(a) := -q(\{x\} \mid x, a)$.

Following the Kitaev construction of a CTMDP [28], we take the sample space $\Omega := S \times ((0, \infty] \times S_\infty)^\infty$, where $S_\infty := S \bigcup \{x_\infty\}$ with the isolated point $x_\infty \notin S$. We equip Ω with its Borel σ-algebra $\mathcal{F}$. For each $n \geq 0$, and any element $\omega := (x_0, \theta_1, x_1, \theta_2, \dots,) \in \Omega$, let $t_n(\omega) := t_{n-1}(\omega) + \theta_n$ with $t_0(\omega) := 0$, and $t_\infty(\omega) := \lim_{n \to \infty} t_n(\omega)$. Obviously, $t_n(\omega)$ are measurable mappings on the sample space Ω. In what follows, we will omit the argument $\omega \in \Omega$ from the presentation for simplicity, and understand t_n, x_n, θ_{n+1}, and t_∞ as the nth jump moment, jumped-in state, holding time of x_n, and the explosion moment, respectively. The pairs $\{t_n, x_n\}$ form a marked point process with the internal history $\{\mathcal{F}_t\}_{t \geq 0}$ (see [27, Chapter 4]), which defines the stochastic process on $(\Omega, \mathcal{F})$ of interest $\{\xi_t,\ t \geq 0\}$ by

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

where x_∞ is the cemetery point so that $A(x_\infty) := \{a_\infty\}$ and $q_{x_\infty}(a_\infty) := 0$ with $a_\infty \notin A$ being some isolated point. Below, we denote $A_\infty := A \cup \{a_\infty\}$. As in [12], we formally put $\xi_\infty := x_\infty$.

Definition 1. A (randomized history-dependent) policy π for the CTMDP is given by a sequence (π_n) such that, for each $n = 0, 1, \dots, \pi_n(\mathrm{d}a \mid x_0, \theta_1, \dots, x_n, s)$ is a stochastic kernel on A concentrated on $A(x_n)$, and for each $\omega = (x_0, \theta_1, x_1, \theta_2, \dots,) \in \Omega$, for $t > 0$,

$$\pi(\mathrm{d}a \mid \omega, t) := \mathbf{1}\{t \geq t_\infty\} \delta_{a_\infty}(\mathrm{d}a) + \sum_{n=0}^{\infty} \mathbf{1}\{t_n < t \leq t_{n+1}\} \pi_n(\mathrm{d}a \mid x_0, \theta_1, \dots, x_n, t - t_n).$$

We can rephrase Definition 1 as follows: a policy π is a predictable (with respect to $\{\mathcal{F}_t\}_{t \geq 0}$) stochastic kernel from $\Omega \times (0, \infty)$ to A_∞; see [27, Theorem 4.19]. The class of all policies for the CTMDP is denoted by Π. A policy is called Markov if it is in the form $\pi(\mathrm{d}a \mid \omega, t) = \pi(\mathrm{d}a \mid \xi_{t-}(\omega), t)$, where, with conventional abuse of notations, π on the right-hand side is a stochastic kernel. Let $\Pi_M \subset \Pi$ be the set of Markov policies.

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

$$\begin{aligned} \nu^\pi(\mathrm{d}t, \mathrm{d}y) &:= \int_A q(\mathrm{d}y \setminus \{\xi_{t-}(\omega)\} \mid \xi_{t-}(\omega), a) \pi(\mathrm{d}a \mid \omega, t)\, \mathrm{d}t \\ &= \sum_{n \geq 0} \int_A q(\mathrm{d}y \setminus \{x_n\} \mid x_n, a) \pi_n(\mathrm{d}a \mid x_0, \theta_1, \dots, x_n, t - t_n)\, \mathbf{1}\{t_n < t \leq t_{n+1}\}\, \mathrm{d}t \end{aligned}$$

with $q(\mathrm{d}y \mid x_\infty, a_\infty) := 0$. Suppose that an initial distribution γ on S is given. Then, by [27, Theorem 4.27], there exists a unique probability measure $\mathbb{P}^\pi_\gamma$ such that

$$\mathbb{P}^\pi_\gamma(\xi_0 \in \mathrm{d}x) = \gamma(\mathrm{d}x),$$

and with respect to $\mathbb{P}^\pi_\gamma$, ν^π is the dual predictable projection of the random measure of the marked point process $\{t_n, x_n\}$. The process $\{\xi_t\}$, defined by (1) under the probability measure $\mathbb{P}^\pi_\gamma$, is called a CTMDP. Below, when $\gamma(\cdot)$ is a Dirac measure concentrated at $x \in S$, we use the denotation $\mathbb{P}^\pi_x$. Expectations with respect to $\mathbb{P}^\pi_\gamma$ and $\mathbb{P}^\pi_x$ are denoted as $\mathbb{E}^\pi_\gamma$ and $\mathbb{E}^\pi_x$,

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

respectively. In fact, in what follows, we often write $\mathbb{P}^\pi$ instead of $\mathbb{P}^\pi_\gamma$ when there is no confusion. Under the probability measure $\mathbb{P}^\pi_\gamma$, the system dynamics of a CTMDP can be described as follows. The initial state x_0 has the distribution given by γ. Given the current state x_n. The sojourn time θ_{n+1} has the conditional tail function given by $\mathbb{P}^\pi(\theta_{n+1} \geq t \mid x_0, \theta_1, \ldots, x_n) = \exp(-\int_0^t \int_A q_{x_n}(a)\pi_n(\mathrm{d}a \mid x_0, \theta_1, \ldots, x_n, s)\,\mathrm{d}s)$, and upon a jump, the conditional distribution of the next state x_{n+1} is given by $\mathbb{P}^\pi(x_{n+1} \in \Gamma \mid x_0, \theta_1, \ldots, x_n, \theta_{n+1}) = (\int_A q(\Gamma \setminus \{x_n\} \mid x_n, a)\pi_n(\mathrm{d}a \mid x_0, \theta_1, \ldots, x_n, \theta_{n+1}))/(\int_A q_{x_n}(a)\pi_n(\mathrm{d}a \mid x_0, \theta_1, \ldots, x_n, \theta_{n+1}))$ for each $\Gamma \in \mathcal{B}(S)$, where, and below, we quite formally put $\int_A q(\Gamma \setminus \{x_n\} \mid a)\pi_n(\mathrm{d}a \mid x_0, \theta_1, \ldots, x_n, \infty) := 0$ for each $\Gamma \in \mathcal{B}(S)$ and use the convention of $\frac{0}{0} := 0$, so that $\mathbb{P}^\pi(x_{n+1} = x_\infty \mid x_0, \theta_1, \ldots, x_n, \theta_{n+1}) = 1 - \mathbb{P}^\pi(x_{n+1} \in S \mid x_0, \theta_1, \ldots, x_n, \theta_{n+1})$. According to [11], under each Markov policy π, the process ξ_t is a Markov jump process in the sense of [12] with respect to $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}^\pi_x)$ for each $x \in S$.

We are also interested in policies in more specific forms.

Definition 2. With slight but conventional abuse of denotations, a policy $\pi = (\pi_n)_{n=0,1,\ldots,} \in \Pi$ is called (randomized) stationary if each of the stochastic kernels π_n reads $\pi_n(\mathrm{d}a \mid x_0, \theta_1, \ldots, x_n, t - t_n) = \pi(\mathrm{d}a \mid x_n)$. A stationary policy is further called deterministic if $\pi_n(\mathrm{d}a \mid x_0, \theta_1, \ldots, x_n, t - t_n) = \delta_{\varphi(x_n)}(\mathrm{d}a)$ for some measurable mapping φ from S to A such that $\varphi(x) \in A(x)$ for each $x \in S$; the existence of such a mapping is guaranteed by the assumption imposed on the multifunction $A(\cdot)$, which also implies the set Π being nonempty.

Let $c(x, a)$, a measurable function on K that takes values in $[0, \infty)$, represent the cost rate at the present state $x \in S$ and action $a \in A(x)$. Quite formally, for any measurable function f on K, we put $f(x_\infty, a_\infty) = 0$. This agreement, together with (1) and that $q(\{x_\infty\} \mid x_\infty, a_\infty) = 0 = q_{x_\infty}(a_\infty)$, allows us to define formally the long-run average cost by

$$W(x, \pi) := \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^\pi_x\left[\int_0^T \int_A c(\xi_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right]$$
$$= \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^\pi_x\left[\int_0^{\min\{T, t_\infty\}} \int_A c(\xi_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right].$$

We are interested in the following optimal control problem

$$W(x, \pi) \to \min_{\pi\in\Pi}, \qquad x \in S, \tag{2}$$

for which a policy π^* is called optimal if $W(x, \pi^*) = \inf_{\pi\in\Pi} W(x, \pi)$ for each $x \in S$.

The objective of this paper is to show the existence of a deterministic stationary optimal policy under the weak continuity conditions on the transition rates, which can be essentially arbitrarily unbounded.

3. Main result

Condition 1. *Let* $\inf_{x\in S} \inf_{\pi\in\Pi} W(x, \pi) < \infty$.

For each real constant $\alpha > 0$, we define the expected total discounted cost under each policy $\pi \in \Pi$ by

$$W_\alpha(x, \pi) := \mathbb{E}^\pi_x\left[\int_0^\infty \mathrm{e}^{-\alpha t} \int_A c(\xi_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right]$$
$$= \mathbb{E}^\pi_x\left[\int_0^{\min\{t_\infty, \infty\}} \mathrm{e}^{-\alpha t} \int_A c(\xi_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right],$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

and the value function for the corresponding discounted problem by $W_\alpha(x) := \inf_{\pi\in\Pi} W_\alpha(x, \pi)$. Let

$$m_\alpha := \inf_{x\in S} W_\alpha(x) \quad \text{and} \quad h_\alpha(x) := W_\alpha(x) - m_\alpha \geq 0,$$

where the regulation of $\infty - \infty := \infty$ is in use. The function h_α on S is sometimes called the relative difference or normalized value function for the discounted problem, on which we impose the following condition, where ρ denotes the predetermined metric on S consistent with its topology.

Condition 2. *We define*

$$\liminf_{0<\alpha\downarrow 0, y\to x} h_\alpha(y) := \sup_{\delta>0,\Delta>0} \left\{ \inf_{0<\alpha\leq\delta, \rho(x,y)<\Delta} h_\alpha(y) \right\} < \infty \quad \textit{for each } x \in S.$$

In Condition 2 we defined the generalized lower limit of the function $h_\alpha(y)$ as $0 < \alpha \downarrow 0$ and $y \to x$. Condition 2 is equivalent to the following; for each $x \in S$, there exist sequences $0 < \alpha_n \downarrow 0$ and $y_n \to x$ such that $\{h_{\alpha_n}(y_n)\}$ is bounded. Condition 2 and its synonyms are widely assumed in the current literature on average CTMDPs. We provide more insight on Condition 2 after we introduce Condition 3, below.

Finally, we assume the following weak continuity condition. To this end, we recall that a function c, on the space K, is called $\mathbb{K}$-inf-compact if it is lower semicontinuous on K, and satisfies the following; for each $S \ni x_n \to x \in S$ as $n \to \infty$, each sequence $a_n \in A(x_n)$ such that $c(x_n, a_n)$ is bounded from the above, admits a limit point $a \in A(x)$ [9]. The function c is called inf-compact on K if the set $\{(x, a) \in K : c(x, a) \leq \lambda\}$ is compact in K for each $\lambda \in (-\infty, \infty)$. By the way, the inf-compactness on K is defined in a weaker sense in [22]. It is known that the inf-compactness of a function implies its $\mathbb{K}$-inf-compactness [9].

Condition 3. (a) *For each bounded continuous function f on S, $\int_S f(y)q(\mathrm{d}y \mid x, a)$ is continuous in $(x, a) \in K$.*

(b) *The cost rate c is $\mathbb{K}$-inf-compact.*

(c) *There exists a continuous function w on S taking values in $(0, \infty)$ such that $\overline{q}_x \leq w(x)$ for each $x \in S$.*

The rather weak part (c) of Condition 3 is for technical convenience; it essentially allows the transition rates to be arbitrarily unbounded, since the function w can also be arbitrarily unbounded. Part (a) of Condition 3 reads that the transition rates are weakly continuous. Condition 3 does not require the multifunction $A(x)$ to be either compact-valued or upper semicontinuous.

Some comments on Condition 2 are in position now. Suppose that Conditions 1 and 3 are satisfied so that, for any $\alpha > 0$, there exists a deterministic stationary optimal policy φ_α^* for the discounted problem, i.e. $W_\alpha(x) = W_\alpha(x, \varphi_\alpha^*)$ for each $x \in S$; and for all sufficiently small $\alpha > 0$, $m_\alpha < \infty$ (as we will explain in the proof of Theorem 1, below). Assume that there exists some $z \in S$ such that, for all sufficiently small $\alpha > 0$, $m_\alpha = W_\alpha(z)$. (In fact, if $S = \{0, 1, 2, \dots, \}$ or $S = [a, b)$ with $a \in \mathbb{R}$ and $b \in \mathbb{R}\cup\{+\infty\}$, then this assumption is satisfied when $A(x)$ is decreasing in $x \in S$, and for all sufficiently small $\alpha > 0$, $c(x, a)/(\alpha + w(x))$ and $(w(x)/(\alpha + w(x))) \int_S u(y)(q(\mathrm{d}y \mid x, a)/w(x) + I\{x \in \mathrm{d}y\})$ are increasing in $x \in S$ for each fixed $a \in A(x)$ and increasing nonnegative function u on S. This follows from the fact that $W_\alpha(x) = \lim_{n\uparrow\infty} v_n(x)$ with v_n being defined in the proof of Lemma 2, below.) Consider the stopping time $\tau_z = \inf\{t \geq 0 : \xi_t = z\}$ (with respect to $\{\mathcal{F}_t\}_{t\geq 0}$). As usual, the infimum taken

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

over the empty set is set as $+\infty$. It is known from [27] that $\min\{\tau_z, t_\infty\}$ is also a stopping time. Then, for all sufficiently small $\alpha > 0$,

$$h_\alpha(x) = \mathbb{E}_x^{\varphi_\alpha^*}\left[\int_0^{\min\{\tau_z,t_\infty\}} \mathrm{e}^{-\alpha t} c(\xi_t, \varphi^*(\xi_t))\,\mathrm{d}t\right] + \mathbb{E}_x^{\varphi_\alpha^*}\left[\mathbb{E}_x^{\varphi_\alpha^*}\left[\int_{\min\{\tau_z,t_\infty\}}^{t_\infty} \mathrm{e}^{-\alpha t} c(\xi_t, \varphi^*(\xi_t))\,\mathrm{d}t \;\middle|\; \mathcal{F}_{\min\{\tau_z,t_\infty\}}\right]\right] - W_\alpha(z).$$

Furthermore, by [12, Theorem 4] the process ξ_t is a strong Markov one with respect to $\{\mathcal{F}_t\}_{t\geq 0}$. So, by applying the strong Markov property to the second summand on the right-hand side of the previous equality, we have

$$\begin{aligned} h_\alpha(x) &\leq \mathbb{E}_x^{\varphi_\alpha^*}\left[\int_0^{\min\{\tau_z,t_\infty\}} \mathrm{e}^{-\alpha t} c(\xi_t, \varphi^*(\xi_t))\,\mathrm{d}t\right] + \mathbb{E}_x^{\varphi_\alpha^*}\left[\mathrm{e}^{-\alpha\min\{\tau_z,t_\infty\}} W_\alpha(z)\right] - W_\alpha(z) \\ &\leq \mathbb{E}_x^{\varphi_\alpha^*}\left[\int_0^{\min\{\tau_z,t_\infty\}} \mathrm{e}^{-\alpha t} c(\xi_t, \varphi_\alpha^*(\xi_t))\,\mathrm{d}t\right] \\ &\leq \sup_\pi \mathbb{E}_x^\pi\left[\int_0^{\min\{\tau_z,t_\infty\}} \int_A c(\xi_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right], \end{aligned}$$

where the first inequality further follows from the fact that ξ_t is right-continuous and $W_\alpha(x_\infty) = 0$. It can be shown [2], [19] that if there is some constant $\varepsilon > 0$ such that $q_x(a) > \varepsilon$ for all $x \neq z$ and $a \in A(x)$, then

$$\sup_\pi \mathbb{E}_x^\pi\left[\int_0^{\min\{\tau_z,t_\infty\}} \int_A c(\xi_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right] < \infty \tag{3}$$

for each $x \in S$ if there exists a real-valued upper semianalytic function v on S such that

$$0 \geq c(x,a) + \int_{S\setminus\{z\}} q(\mathrm{d}y \mid x, a) v(y)$$

for each $x \neq z$ and $a \in A(x)$. This provides a sufficient condition imposed on the primitives of the CTMDP model for verifying Condition 2, which does not refer to the existence of a Lyapunov function as in Condition 4, below (cf. [18]). By the way, if the process ξ_t is nonexplosive, as prevailingly assumed in the current literature, then (3) is satisfied when, for example, the process ξ_t exhibits some version of the ergodic property.

Similar versions of Conditions 1, 2 and parts (a) and (b) of Condition 3 are assumed in [8] but for discrete-time problems; see Assumptions G, W* and B therein.

Theorem 1. *Suppose that Conditions 1, 2, and 3 are satisfied. Then there exist a constant g, a nonnegative real-valued lower semicontinuous function h on S, and a deterministic stationary policy φ^* such that*

(a) *the following optimality inequality is satisfied for each $x \in S$:*

$$\begin{aligned} g + w(x)h(x) &\geq \inf_{a\in A(x)}\left\{c(x,a) + w(x)\int_S h(y)\left(\frac{q(\mathrm{d}y \mid x, a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\}\right)\right\} \\ &= c(x, \varphi^*(x)) + w(x)\int_S h(y)\left(\frac{q(\mathrm{d}y \mid x, \varphi^*(x))}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\}\right); \end{aligned} \tag{4}$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

(b) *the deterministic stationary policy φ^* is optimal for the average CTMDP problem (2); and*

(c) $g = \inf_{\pi\in\Pi} W(x,\pi) < \infty$ *for each* $x \in S$.

The proof of Theorem 1 is given in Section 4, by inspecting which we can see that any deterministic stationary policy that satisfies (4) is optimal. Furthermore, it follows from Lemma 1, below, that $\inf_\pi W(x,\pi)$ is given by the smallest constant g satisfying the inequality (4).

The statement of Theorem 1 is obtained in [18] under the following condition; see Assumptions A, B, and C therein.

Condition 4. (a) *There exists a measurable function $w \geq 1$ on S and constants $c_0 \in (-\infty, \infty)$, $b_0 \geq 0$, and $M_0 > 0$ such that*

(i) $\int_S w(y)q(\mathrm{d}y \mid x,a) \leq c_0 w(x) + b_0$ *for each* $(x,a) \in K$; *and*

(ii) $\overline{q}_x \leq M_0 w(x)$ *for all* $x \in S$.

(b) *For some sequence $\alpha_n \downarrow 0$ as $n \uparrow \infty$ and some fixed $x_0 \in S$, there exist a real constant L^* and a finitely valued nonnegative measurable function U on S such that*

(i) $\sup_{n=1,2,\ldots}\{\alpha_n W_{\alpha_n}(x)\} < \infty$ *for each* $x \in S$; *and*

(ii) $L^* \leq W_{\alpha_n}(x) - W_{\alpha_n}(x_0) \leq U(x) < \infty$ *for each* $x \in S$.

(c) *The following compactness-continuity condition is satisfied:*

(i) *the set $A(x)$ is compact for each $x \in S$;*

(ii) *the cost rate $c(x,a)$ is lower semicontinuous in $a \in A(x)$ for each $x \in S$; and*

(iii) *for each bounded measurable function f on S, $\int_S f(y)q(\mathrm{d}y \mid x,a)$ is continuous in $a \in A(x)$ for each $x \in S$.*

The function w in part (a) of Condition 4 is called a Lyapunov function or a bounding function, whose existence guarantees the process ξ_t to be nonexplosive, i.e. $\mathbb{P}_x^\pi(t_\infty = \infty) = 1$ for each $x \in S$ [18], which is also prevailingly assumed in the previous literature on CTMDPs with possibly unbounded transition rates [13], [14], [15], [17], [20], [21], [30], [32], [34]. In comparison, the existence of a Lyapunov function is not needed in the present paper; Condition 3(c) allows essentially arbitrarily unbounded transition rates, and thus, the underlying process to be explosive. Condition 4(c)(iii) states the strong continuity of $q(\mathrm{d}y \mid x,a)$; accordingly, the lower semicontinuity of the cost rate $c(x,a)$ is only required in $a \in A(x)$, but the multifunction $A(\cdot)$ needs to be compact-valued, which is not required in this paper. Finally, we note that Condition 4 implies Condition 1.

4. Proof of Theorem 1

In this section, before proving Theorem 1, we first present some auxiliary statements. Under each Markov policy $\pi \in \Pi_M$, the process ξ_t is a Markov jump process [11], and there exists a transition (subprobability, in general) function $p^\pi(u,x,t,\mathrm{d}y)$ such that $\mathbb{P}^{p^\pi}(\xi_t \in \mathrm{d}y \mid \xi_u) = p^\pi(u,\xi_u,t,\mathrm{d}y)$ with $t \geq u \geq 0$ almost surely with respect to $\mathbb{P}^{p^\pi}$ [29]. So we formally define for each $x \in S$, with $u \leq t$ and Markov policy $\pi \in \Pi_M$

$$W^\pi(u,x,t) := \int_u^t \int_S \int_A c(y,a)\pi(\mathrm{d}a \mid y,s)p^\pi(u,x,s,\mathrm{d}y)\,\mathrm{d}s. \tag{5}$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

The next result is a generalization of [18, Theorem 3.4], which was proved for deterministic stationary policies only and additionally under Condition 4(a). Since Condition 4(a) is not required in this paper, to be self-contained and for its potential independent interest, we include this result here, and present its complete proof in Appendix A.

Lemma 1. (a) *Let a Markov policy $\pi \in \Pi_M$ be fixed. Then the function $W^\pi(u,x,t)$ is the minimal nonnegative measurable solution to the following inequality*

$$\begin{aligned} v(u,x,t) \ge & \int_u^t \int_A c(x,a)\pi(\mathrm{d}a \mid x,\theta)\,\mathrm{d}\theta \exp\biggl(-\int_u^t \int_A q_x(a)\pi(\mathrm{d}a \mid x,\theta)\,\mathrm{d}\theta\biggr) \\ & + \int_u^t \exp\biggl(-\int_u^s \int_A q_x(a)\pi(\mathrm{d}a \mid x,\theta)\,\mathrm{d}\theta\biggr) \\ & \quad \times \biggl\{\int_A q_x(a)\pi(\mathrm{d}a \mid x,s) \int_u^s \int_A c(x,a)\pi(\mathrm{d}a \mid x,\theta)\,\mathrm{d}\theta \\ & \qquad + \int_{S\setminus\{x\}} \int_A q(\mathrm{d}y \mid x,a)\pi(\mathrm{d}a \mid x,s) v(s,y,t)\biggr\}\,\mathrm{d}s. \end{aligned} \tag{6}$$

(b) *Let a stationary policy π be fixed, and suppose there exist a constant $g \in [0,\infty]$ and a nonnegative measurable function h on S satisfying the following inequality*

$$g + h(x)\int_A q_x(a)\pi(\mathrm{d}a \mid x) \ge \int_A c(x,a)\pi(\mathrm{d}a \mid x) + \int_{S\setminus\{x\}} h(y) \int_A q(\mathrm{d}y \mid x,a)\pi(\mathrm{d}a \mid x)$$

for each $x \in S$. Then $g \ge W(x,\pi)$ for each $x \in S$ such that $h(x) < \infty$.

Proof. See Appendix A.

The next lemma, to be used in the proof of Theorem 1, below, extends some known results for discounted CTMDPs in the literature [6], [13] to weaker conditions.

Lemma 2. *Suppose that Condition 3(c) is satisfied. For each $\alpha > 0$, W_α is the minimal nonnegative lower semi-analytic solution to the equation*

$$v(x) = \inf_{a\in A(x)} \biggl\{\frac{c(x,a)}{\alpha + w(x)} + \frac{w(x)}{w(x)+\alpha} \int_S v(y)\biggl(\frac{q(\mathrm{d}y \mid x,a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\}\biggr)\biggr\}. \tag{7}$$

If, additionally, Condition 3(a) and (b) also hold, then W_α is lower semicontinuous on S, and there exists a deterministic stationary optimal policy for the discounted CTMDP problem.

Proof. See Appendix A.

Proof of Theorem 1. Note that under Condition 1, $m_\alpha < \infty$ by [15, Proposition A.5] for all sufficiently small $\alpha > 0$, say, to be specific, for all $0 < \alpha \le \alpha_0 < \infty$. Indeed, Condition 1 asserts the existence of some $z \in S$ and policy $\pi \in \Pi$ such that $W(z,\pi) = \limsup_{t\uparrow\infty}(1/t)\mathbb{E}_z^\pi[\int_0^t \int_A c(\xi_s,a)\pi(\mathrm{d}a \mid \omega,s)\,\mathrm{d}s] < \infty$. Thus, for all sufficiently large $t > 0$, $\int_0^t \mathbb{E}_z^\pi[\int_A c(\xi_s,a)\pi(\mathrm{d}a \mid \omega,s)]\,\mathrm{d}s = \mathbb{E}_z^\pi[\int_0^t \int_A c(\xi_s,a)\pi(\mathrm{d}a \mid \omega,s)\,\mathrm{d}s] < \infty$. Due to the nonnegativity of the cost rate c, this implies $\mathbb{E}_z^\pi[\int_A c(\xi_t,a)\pi(\mathrm{d}a \mid \omega,t)] < \infty$ for $t > 0$ almost everywhere. Thus, the condition of Proposition A.5 in [15] is verified, and from it, we infer that

$$\limsup_{0<\alpha\downarrow 0} \alpha W_\alpha(z,\pi) \le W(z,\pi) < \infty, \tag{8}$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

and consequently, there exists some $0 < \alpha_0 < \infty$ such that $m_\alpha \le W_\alpha(z,\pi) < \infty$ for all $0 < \alpha \le \alpha_0$ as required.

Let $g_\alpha := \alpha m_\alpha$. For each $0 < \alpha \le \alpha_0$, we write $W_\alpha(x) = h_\alpha(x) + m_\alpha$ in (7) with W_α in lieu of v, and obtain

$$\begin{aligned} h_\alpha(x) + m_\alpha &= \inf_{a\in A(x)} \left\{ \frac{c(x,a)}{\alpha + w(x)} \right. \\ &\qquad \left. + \frac{w(x)}{w(x)+\alpha} \int_S (h_\alpha(y) + m_\alpha) \left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \right\} \\ &= \inf_{a\in A(x)} \left\{ \frac{c(x,a)}{\alpha + w(x)} + \frac{w(x)}{w(x)+\alpha} \int_S h_\alpha(y) \left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \right. \\ &\qquad \left. + \frac{w(x) m_\alpha}{\alpha + w(x)} \right\}. \end{aligned} \tag{9}$$

It follows from (9) that

$$\begin{aligned} (w(x)+\alpha) h_\alpha(x) + g_\alpha &= \inf_{a\in A(x)} \left\{ c(x,a) + w(x) \right. \\ &\qquad \left. \times \int_S h_\alpha(y) \left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \right\}. \end{aligned} \tag{10}$$

Now define

$$g := \limsup_{0<\alpha\downarrow 0} g_\alpha \ge 0, \tag{11}$$

which is finite because of (8), and

$$h(x) := \liminf_{0<\alpha\downarrow 0,\, y\to x} h_\alpha(y), \tag{12}$$

which is finite under Condition 2. It is known that for each convergent sequence $0 < \alpha_n \downarrow 0$ as $n \to \infty$,

$$\sup_{\alpha\in(0,\infty)} \underline{h}_\alpha(x) = h(x) = \liminf_{n\to\infty,\, y\to x} \underline{h}_{\alpha_n}(y), \tag{13}$$

where $\underline{h}_{\alpha_n}(x) := \liminf_{y\to x} H_{\alpha_n}(y)$ with $H_{\alpha_n}(y) := \inf_{\alpha\in(0,\alpha_n]} h_\alpha(y)$; see [8, Equation (24)] for the first equality in (13) and [8, Corollary 1] for the other. The above three functions are all measurable; in fact, the functions $\underline{h}_\alpha$ and h are lower semicontinuous on S; see [1, Lemma 5.13.4] and [4, Lemma 4.2], respectively. Note that, by their definitions

$$h_\beta(x) \ge H_\beta(x) \ge H_\alpha(x) \ge \underline{h}_\alpha(x) \tag{14}$$

for each $x \in S$ and $\alpha \ge \beta > 0$.

Let $\varepsilon > 0$ be arbitrarily fixed. Then, by the definition of the constant g (see (11)), there exists $0 < \alpha_1 \le \alpha_0$ such that, for each $\alpha \in (0, \alpha_1]$, $g \ge g_\alpha - \varepsilon$. It follows from this, and (10), that for each $0 < \alpha \le \alpha_1$,

$$\begin{aligned} &(w(x)+\alpha) h_\alpha(x) + g + \varepsilon \\ &\quad \ge \inf_{a\in A(x)} \left\{ c(x,a) + w(x) \int_S h_\alpha(y) \left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \right\}, \end{aligned}$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

which, together with (14), leads to, for each $0 < \beta \le \alpha \le \alpha_1$,

$$(w(x)+\beta)h_\beta(x)+g+\varepsilon \ge \inf_{a\in A(x)}\left\{c(x,a)+w(x)\int_S H_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)\right\};$$

thus, by the definition of H_α, the above relation, and (14), again provide

$$\begin{aligned}(w(x)+\alpha)H_\alpha(x)+g+\varepsilon &\ge \inf_{a\in A(x)}\left\{c(x,a)+w(x)\int_S H_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)\right\}\\ &\ge \inf_{a\in A(x)}\left\{c(x,a)+w(x)\int_S \underline{h}_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)\right\}, \qquad (15)\end{aligned}$$

for each $0 < \alpha \le \alpha_1$. Under Condition 3, the stochastic kernel $q(\mathrm{d}y \mid x,a)/w(x)+\mathbf{1}\{x\in \mathrm{d}y\}$ is weakly continuous, which, together with the lower semicontinuity of $\underline{h}_\alpha$ (as explained earlier), implies that

$$w(x)\int_S \underline{h}_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)$$

defines a lower semicontinuous function on S. As a result,

$$c(x,a)+w(x)\int_S \underline{h}_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)$$

is $\mathbb{K}$-inf-compact because so is the cost rate c and that $\underline{h}_\alpha(x) \ge 0$ for each $x \in S$; see Lemma 4, below. Therefore, we can infer from Lemma 3, below, for the lower semicontinuity on S of the expression in the second line of (15). Following from this and upon taking the corresponding lower limit on the both sides of (15), we obtain

$$(w(x)+\alpha)\underline{h}_\alpha(x)+g+\varepsilon \ge \inf_{a\in A(x)}\left\{c(x,a)+w(x)\int_S \underline{h}_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)\right\}$$

for each $0 < \alpha \le \alpha_1$. The first equality of (13) and the above inequality imply

$$(w(x)+\alpha)h(x)+g+\varepsilon \ge \inf_{a\in A(x)}\left\{c(x,a)+w(x)\int_S \underline{h}_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)\right\} \qquad (16)$$

for each $0 < \alpha \le \alpha_1$. By the $\mathbb{K}$-inf-compactness of the expression inside the parenthesis on the right-hand side of (16) (as explained earlier) and Lemma 3, for each $0 < \alpha \le \alpha_1$, there exists some $a_\alpha \in A(x)$ such that

$$\begin{aligned}(w(x)+\alpha)h(x)+g+\varepsilon &\ge \inf_{a\in A(x)}\left\{c(x,a)+w(x)\int_S \underline{h}_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right)\right\}\\ &= c(x,a_\alpha)+w(x)\int_S \underline{h}_\alpha(y)\left(\frac{q(\mathrm{d}y \mid x,a_\alpha)}{w(x)}+\mathbf{1}\{x\in \mathrm{d}y\}\right). \qquad (17)\end{aligned}$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

Now let $x \in S$ be arbitrarily fixed, and take $\alpha_1 \geq \alpha_n \downarrow 0$ ($\alpha_n > 0$). Under Condition 2 the expression on the left-hand side of (17) is finite (recall (12) for the definition of the function h). Considering (17) with α_n replacing α therein, it follows from the definition of the $\mathbb{K}$-inf-compactness that the sequence $\{a_{\alpha_n}\}$ admits a limit point $a^* \in A(x)$. Taking the lower limit on the both sides of (17) along the specified sequence $\alpha_1 \geq \alpha_n \downarrow 0$ ($\alpha_n > 0$), we have

$$\begin{aligned} & w(x)h(x) + g + \varepsilon \\ & \quad \geq c(x, a^*) + w(x) \liminf_{n\to\infty} \int_S \underline{h}_{\alpha_n}(y) \left(\frac{q(\mathrm{d}y \mid x, a_{\alpha_n})}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \\ & \quad \geq c(x, a^*) + w(x) \int_S h(y) \left(\frac{q(\mathrm{d}y \mid x, a^*)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \\ & \quad \geq \inf_{a \in A(x)} \left\{ c(x, a) + w(x) \int_S h(y) \left(\frac{q(\mathrm{d}y \mid x, a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \right\}, \end{aligned} \tag{18}$$

where for the first inequality the finiteness of $h(x)$ and the lower semicontinuity of the term inside the parenthesis on the right-hand side of (16) are used; and the second inequality follows from (13), the weak continuity of the underlying stochastic kernel, and the generalized Fatou's lemma; see Lemma 5, below, or [4, Lemma 4.2]. That the inequality in (4) is satisfied by the constant g and the nonnegative real-valued lower semicontinuous function h follows from (18) and the arbitrariness of $\varepsilon > 0$. Regarding the existence of a measurable selector φ^* satisfying the equality in (4), we can refer to Lemma 3; recall that the term in the parenthesis in (4) is $\mathbb{K}$-inf-compact. We prove the rest of this statement as follows. Let φ^* be any measurable selector satisfying the equality in (4). By the finiteness of $h(x)$, (4), and Lemma 1, we have

$$g \geq W(x, \varphi^*) \geq \inf_{\pi \in \Pi} W(x, \pi). \tag{19}$$

For the opposite direction, let $x \in S$ be arbitrarily fixed. Since $g < \infty$, we see that $\inf_{\pi\in\Pi} W(x, \pi) < \infty$. Fix arbitrarily some (possibly x-dependent) policy π such that $W(x, \pi) < \infty$. Now, as in the argument for (8) with z being replaced by x in the beginning of this proof, we see $\limsup_{0<\alpha\downarrow 0} \alpha W_\alpha(x) \leq W(x, \pi) < \infty$, which together with the arbitrariness of the policy π and the fact that $g = \limsup_{0<\alpha\downarrow 0} \alpha \inf_{x\in S} W_\alpha(x) \leq \limsup_{0<\alpha\downarrow 0} \alpha W_\alpha(x)$ (recalling here the definition of g given by (11)), leads to $\inf_{\pi\in\Pi} W(x, \pi) \geq g$. Thus, we see the validity of (19) with inequalities being replaced by equalities. It follows from the arbitrariness of $x \in S$ that the policy φ^* is optimal. The proof is now completed.

5. Conclusion

To sum up, for a CTMDP in Borel state and action spaces with a nonnegative cost rate, the existence of a deterministic stationary average optimal policy is proved with weakly continuous transition rates. Our conditions allow the controlled process to be explosive (i.e. the transition rates are essentially arbitrarily unbounded). In addition, following the neat generalization of the Berge theorem [9], the condition on the admissible action spaces has been further relaxed as compared with the previous literature.

Appendix A.

Definition 3. The collection of analytic subsets of a nonempty Borel space S is the collection of images of measurable subsets of Y under all measurable mappings from Y into S, where Y

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

is an uncountable Borel space. A function f on the nonempty Borel space S is called lower semianalytic if for each $\varepsilon \in (-\infty, \infty)$, the set $\{x \in S\colon f(x) < \varepsilon\}$ is analytic. A function f is called upper semianalytic if $-f$ is lower semianalytic.

We refer the reader to [2, Chapter 7] for more details on the above definition.

The next lemma comes from [9, Theorems 1.2 and 3.3], where more general statements are established.

Lemma 3. *Suppose that a function g on the nonempty Borel space $K = \{(x, a)\colon x \in S,\ a \in A(x)\}$ is $\mathbb{K}$-inf-compact. Then $\inf_{a\in A(x)} g(x, a)$ defines a lower semicontinuous function in $x \in S$. Furthermore, there is a measurable mapping φ^* from S to A, whose graph is contained in K, such that $\inf_{a\in A(x)} g(x, a) = g(x, \varphi^*(x))$ for each $x \in S$.*

The following lemma summarizes some facts about $\mathbb{K}$-inf-compact functions, which are used frequently in the proofs in this paper.

Lemma 4. *Let c be a $\mathbb{K}$-inf-compact function on K. If v is a nonnegative lower semicontinuous function on K, then $c + v$ is also $\mathbb{K}$-inf-compact on K. If u is a continuous real-valued function on S such that $u(x) > 0$ for each $x \in S$, then $c(x, a)/u(x)$ defines a $\mathbb{K}$-inf-compact function on K.*

Proof. We only verify the second part. Clearly $c(x, a)/u(x)$ is lower semicontinuous on K. Now suppose that $S \ni x_n \to x \in S$ and $a_n \in A(x_n)$ such that there is some real constant $M > 0$ such that $c(x_n, a_n)/u(x_n) \le M$, i.e. $c(x_n, a_n) \le Mu(x_n)$ for all n. Since u is continuous and the set $X := \bigcup_{n=0}^{\infty}\{x_n\} \cup \{x\}$ is compact in S, we further infer from the previous inequality for that $c(x_n, a_n) \le M \sup_{y\in X} u(y) < \infty$ for all n. Now it follows from the $\mathbb{K}$-inf-compactness of the function c that there exists a limit point $a \in A(x)$ for the sequence $\{a_n\}$, as required.

The following statement is known as the generalized Fatou's lemma [4], [5], [10]. A detailed proof with more general statements is available in [10].

Lemma 5. *Suppose that a sequence of probability measures Q_n on the nonempty Borel space $\mathcal{B}(S)$ is weakly convergent to the probability measure Q on $\mathcal{B}(S)$. Then for each sequence of nonnegative functions g_n on S, it holds that*

$$\int_S \Big(\liminf_{n\to\infty, x\to y} g_n(x)\Big) Q(\mathrm{d}y) \le \liminf_{n\to\infty} \int_S g_n(y)\, Q_n(\mathrm{d}y).$$

Proof of Lemma 1. (a) For simplicity, throughout the proof of this lemma, we omit the fixed policy π from indications, and introduce the following notation:

$$\begin{aligned} c(x, s) &:= \int_A c(x, a)\pi(\mathrm{d}a \mid x, s), \\ q_x(s) &:= \int_A q_x(a)\pi(\mathrm{d}a \mid x, s), \\ q(\mathrm{d}y \mid x, s) &:= \int_A q(\mathrm{d}y \mid x, a)\pi(\mathrm{d}a \mid x, s). \end{aligned}$$

Furthermore, if $c(x, s)$, $q_x(s)$, and $q(\mathrm{d}y \mid x, s)$ in the above are s-independent, as in the case of a stationary policy, we omit s from the arguments.

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

It is known from [11] that the transition function $p(u, x, t, \mathrm{d}y)$ can be constructed iteratively by $\sum_{k=0}^{n} p_k(u, x, t, \mathrm{d}y) \uparrow p(u, x, t, \mathrm{d}y)$ as $n \uparrow \infty$, where the convergence is set-wise, and for each $\Gamma \in \mathcal{B}(S)$

$$p_0(u, x, t, \Gamma) := \mathbf{1}\{x \in \Gamma\} \exp\left(-\int_u^t q_x(s)\,\mathrm{d}s\right);$$
$$p_k(u, x, t, \Gamma) := \int_u^t \int_{S\setminus\{x\}} \exp\left(-\int_u^s q_x(\theta)\,\mathrm{d}\theta\right) q(\mathrm{d}y \mid x, s)\, p_{k-1}(s, y, t, \Gamma)\,\mathrm{d}s.$$

It follows from this, the nonnegativity of the cost rate c, and the monotone convergence theorem; see [23, Theorem 2.1], that $m_n(u, x, t) := \int_u^t \int_S c(y, s) \sum_{k=0}^{n} p_n(u, x, s, \mathrm{d}y)\,\mathrm{d}s \uparrow W(u, x, t)$ as $n \uparrow \infty$; see (5).

We first verify that $W(u, x, t)$ satisfies (6) with the equality as follows. By the iterative definitions of the transition functions p_n,

$$\begin{aligned}
m_n(u, x, t) &:= \int_u^t \int_S c(y, s) \sum_{k=0}^{n} p_k(u, x, s, \mathrm{d}y)\,\mathrm{d}s \\
&= m_0(u, x, t) + \int_u^t \int_S c(y, s) \sum_{k=1}^{n} p_k(u, x, s, \mathrm{d}y)\,\mathrm{d}s \\
&= m_0(u, x, t) + \int_u^t \int_S c(y, s) \Bigg[\sum_{k=1}^{n} \int_u^s \int_{S\setminus\{x\}} \exp\left(-\int_u^r q_x(\theta)\,\mathrm{d}\theta\right) \\
&\qquad\qquad \times q(\mathrm{d}z \mid x, r) \\
&\qquad\qquad \times p_{k-1}(r, z, s, \mathrm{d}y)\,\mathrm{d}r\,\mathrm{d}s\Bigg] \\
&= m_0(u, x, t) + \int_u^t \int_S c(y, s) \Bigg[\sum_{k-1=0}^{n-1} \int_r^t \int_{S\setminus\{x\}} \exp\left(-\int_u^r q_x(\theta)\,\mathrm{d}\theta\right) \\
&\qquad\qquad \times q(\mathrm{d}z \mid x, r) \\
&\qquad\qquad \times p_{k-1}(r, z, s, \mathrm{d}y)\,\mathrm{d}s\,\mathrm{d}r\Bigg] \\
&= m_0(u, x, t) + \int_u^t \exp\left(-\int_u^r q_x(\theta)\,\mathrm{d}\theta\right) \int_{S\setminus\{x\}} q(\mathrm{d}z \mid x, r) m_{n-1}(r, z, t)\,\mathrm{d}r,
\end{aligned}$$

where the last two inequalities follow from the legal interchange of the order of integrations. Integration by parts gives

$$\begin{aligned}
m_0(u, x, t) &= \exp\left(-\int_u^t q_x(\theta)\,\mathrm{d}\theta\right) \int_u^t c(x, \theta)\,\mathrm{d}\theta \\
&\quad + \int_u^t \int_u^s c(x, \theta)\,\mathrm{d}\theta \exp\left(-\int_u^s q_x(\theta)\,\mathrm{d}\theta\right) q_x(s)\,\mathrm{d}s.
\end{aligned}$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

Thus, it follows that

$$m_n(u,x,t) = \int_u^t c(x,\theta)\,\mathrm{d}\theta \exp\left(-\int_u^t q_x(\theta)\,\mathrm{d}\theta\right) + \int_u^t \exp\left(-\int_u^s q_x(\theta)\,\mathrm{d}\theta\right) \times \left\{q_x(s)\int_u^s c(x,\theta)\,\mathrm{d}\theta + \int_{S\setminus\{x\}} q(\mathrm{d}y \mid x,s) m_{n-1}(s,y,t)\right\}\mathrm{d}s. \quad (20)$$

By the standard monotone convergence theorem, passing to the limit as $n \uparrow \infty$ on both sides of the above equality gives

$$W(u,x,t) = \int_u^t c(x,\theta)\,\mathrm{d}\theta \exp\left(-\int_u^t q_x(\theta)\,\mathrm{d}\theta\right) + \int_u^t \exp\left(-\int_u^s q_x(\theta)\,\mathrm{d}\theta\right) \times \left\{q_x(s)\int_u^s c(x,\theta)\,\mathrm{d}\theta + \int_{S\setminus\{x\}} q(\mathrm{d}y \mid x,s) W(s,y,t)\right\}\mathrm{d}s.$$

For the minimality of $W(u,x,t)$ as a nonnegative measurable solution to (6), suppose that there is another nonnegative measurable solution $v(u,x,t)$ to (6). Thus, $v(u,x,t) \geq m_0(u,x,t)$. Now an inductive argument based on (20) and the fact that v satisfies (6) implies $v(u,x,t) \geq m_n(u,x,t)$ for each $n = 0, 1, \ldots$, which, together with the fact that $m_n \uparrow W$ pointwise as $n \uparrow \infty$, leads to $v(u,x,t) \geq W(u,x,t)$ as desired.

(b) Suppose that a stationary policy π is fixed, and there exist a constant g and a nonnegative measurable function h on S as in the statement. Without loss of generality, we assume that $g < \infty$, otherwise the statement holds automatically. It is well known, or otherwise follows from the construction of the transition function $p(u,x,t,\mathrm{d}y)$ above, that under the stationary policy, $p(u,x,t,\mathrm{d}y)$ depends on u and t only through the time increment $t-u$, and the underlying Markov jump process ξ_t is homogeneous, and thus $W(u,x,t) = \mathbb{E}_x[\int_0^{t-u} c(\xi_s)\,\mathrm{d}s] =: \tilde{W}(x,t-u)$; see [11, Theorem 2.2]; recall the agreement that the (stationary) policy π is omitted from indication in this proof. It follows from this and part (a) specialized to a stationary policy and $u = 0$, that $\tilde{W}(x,t)$ is the minimal nonnegative measurable solution to the inequality

$$\tilde{W}(x,t) \geq c(x)t\mathrm{e}^{-q_x t} + \int_0^t \mathrm{e}^{-q_x s}\left\{q_x c(x)s + \int_{S\setminus\{x\}} q(\mathrm{d}y \mid x)\tilde{W}(y,t-s)\right\}\mathrm{d}s.$$

Now it can be verified, based on the definitions of the constant g and the function h, that the above inequality is satisfied with $h(x) + gt$ in lieu of $\tilde{W}(x,t)$. Consequently, $h(x) + gt \geq \tilde{W}(x,t)$ by part (a) of this lemma. At $x \in S$ such that $h(x) < \infty$, dividing the both sides of the previous inequality and then passing to the upper limit as $t \to \infty$ yields the statement.

Proof of Lemma 2. Let $\alpha > 0$ be arbitrarily fixed. It is known that the value function W_α for the discounted CTMDP problem is the minimal nonnegative lower semianalytic solution to the equation

$$v(x) = \inf_{a\in A(x)}\left\{\frac{c(x,a)}{\alpha + q_x(a)} + \int_{S\setminus\{x\}} v(y)\frac{q(\mathrm{d}y \mid x,a)}{\alpha + q_x(a)}\right\} =: \tilde{T}\circ v(x); \quad (21)$$

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

see [6, Theorem 5.5.5]. For the first part of this lemma, it remains to recognize that the two equations (7) and (21) admit the same minimal nonnegative solution. Below, in spite of the argument being trivial, we briefly verify this relation because first, a similar relation between (7) and another equation similar to (21) was falsely claimed without proofs in [31] (see [31, Equation (8)]), and second, it is easy to construct examples to show that equations (7) and (21) are not equivalent; indeed, there can be solutions to (7), which do not satisfy (21). For brevity, we write (7) as $v = T \circ v$ with $T \circ v(x) := \inf_{a \in A(x)}\{((c(x,a)/(\alpha + w(x))) + (w(x)/(w(x)+\alpha)) \int_S v(y)(((q(\mathrm{d}y \mid x,a))/w(x)) + \mathbf{1}\{x \in \mathrm{d}y\})\}$.

Firstly, consider the minimal nonnegative solution u to (21), and let $x \in S$ be arbitrarily fixed. If $u(x) = \infty$, then $T \circ u(x) = \infty = u(x)$ (recalling the convention of $\infty - \infty := \infty$). Now suppose that $u(x) < \infty$. Then it follows that $u(x) \le (c(x,a)/(\alpha + w(x))) + (w(x)/(w(x)+\alpha)) \int_S u(y)((q(\mathrm{d}y \mid x,a)/w(x)) + \mathbf{1}\{x \in \mathrm{d}y\})$ for each $a \in A(x)$. Let $\delta > 0$ be arbitrarily fixed, and take any $0 < \varepsilon < \delta$. Then there exists some $a_\delta \in A(x)$ such that $u(x) + \varepsilon \ge (c(x,a_\delta)/(\alpha + q_x(a_\delta))) + \int_{S\setminus\{x\}} u(y)(q(\mathrm{d}y \mid x,a_\delta)/(\alpha + q_x(a_\delta)))$ so that $u(x) + \delta > u(x) + (\varepsilon(\alpha + q_x(a_\delta))/(\alpha + w(x))) \ge (c(x,a_\delta)/(\alpha + w(x))) + (w(x)/(w(x) + \alpha)) \int_S u(y)((q(\mathrm{d}y \mid x,a_\delta)/w(x)) + \mathbf{1}\{x \in \mathrm{d}y\})$. Since $\delta > 0$ is arbitrarily fixed, we see that $u(x) = T \circ u(x)$. Thus, $u \ge v$ with v being the minimal nonnegative solution to (7). For the opposite direction, note that if $v(x) = \infty$, then $v(x) \ge \tilde{T} \circ v(x)$. Now suppose that $v(x) < \infty$. Then for each $a \in A(x)$, $v(x) \le (c(x,a)/(\alpha + w(x))) + (w(x)/(w(x) + \alpha)) \int_S v(y)((q(\mathrm{d}y \mid x,a)/w(x)) + \mathbf{1}\{x \in \mathrm{d}y\})$, and so $v(x) \le (c(x,a)/(\alpha + q_x(a))) + \int_{S\setminus\{x\}} v(y)(q(\mathrm{d}y \mid x,a)/(q_x(a) + \alpha))$.

Let $\delta > 0$ be arbitrarily fixed, and choose $\varepsilon > 0$ such that $\varepsilon(\alpha + w(x))/\alpha < \delta$. Since v satisfies (7), there exists some $a_\delta \in A(x)$ such that $v(x) \ge (c(x,a_\delta)/(\alpha + w(x))) + (1/(\alpha + w(x))) \int_S v(y) q(\mathrm{d}y \mid x,a_\delta) + (w(x)v(x)/(\alpha + w(x))) - \varepsilon$. Simple rearrangements of this inequality further lead to $v(x) \ge (c(x,a_\delta)/(\alpha + q_x(a_\delta))) + (1/(\alpha + q_x(a_\delta))) \int_{S\setminus\{x\}} v(y) q \times (\mathrm{d}y \mid x,a_\delta) - \delta$. Thus, $v(x) \ge \tilde{T} \circ v(x)$. It follows from this and [2, Proposition 9.10] that $u \le v$; thus, $u = v$ (recalling the opposite direction of the previous inequality being established earlier). The first part of this lemma is proved.

Next, we observe that according to the first part of this lemma and [2, Proposition 9.16], W_α is also given by the value function of a DTMDP with the total undiscounted cost criterion specified by the following primitives. The state space is $S \cup \{x_\infty\}$; the action space is $A \cup \{a_\infty\}$; the admissible action space is $A(x)$ for each $x \in S$ with $A(x_\infty) = \{a_\infty\}$; the transition probability is given by $Q(\Gamma \mid x,a) := (w(x)/(w(x)+\alpha))((q(\Gamma \mid x,a)/w(x)) + \mathbf{1}\{x \in \Gamma\})$ for each $x \in S$, $a \in A(x)$, and $\Gamma \in \mathcal{B}(S)$, $Q(\{x_\infty\} \mid x,a) := \mathbf{1}\{x = x_\infty,\ a = a_\infty\} + \mathbf{1}\{x \in S,\ a \in A(x)\}(1 - Q(S \mid x,a))$; and finally, the cost function is $\mathbf{1}\{x \in S,\ a \in A(x)\}(c(x,a)/(\alpha + w(x)))$. Here we recall that $x_\infty \notin S$ and $a_\infty \notin A$ are two isolated points. Under Condition 3, we can verify that the transition probability $Q(\mathrm{d}y \mid x,a)$ is weakly continuous, i.e. for each bounded continuous function f on $S \cup \{x_\infty\}$, $\int_{S\cup\{x_\infty\}} f(y) Q(\mathrm{d}y \mid x,a)$ is continuous in $x \in S \cup \{x_\infty\}$ and $a \in A(x)$; and the cost function is $\mathbb{K}$-inf-compact; see Lemma 4. Denote the value function for this DTMDP problem with the total undiscounted cost criterion also by W_α. Below, to be self-contained, we verify that W_α can be constructed using the value iteration algorithm under Condition 3. Let $v_0(x) := 0$ and $v_n(x) := \inf_{a \in A(x)}\{(c(x,a)/(\alpha + w(x))) + (w(x)/(w(x)+\alpha)) \int_S v_{n-1}(y)((q(\mathrm{d}y \mid x,a)/w(x)) + \mathbf{1}\{x \in \mathrm{d}y\})\}$ for each $x \in S$, whereas $v_n(x_\infty) := 0$ for each $n = 0, 1, 2, \ldots$. Under Condition 3, since the transition probability is weakly continuous and the cost function is $\mathbb{K}$-inf-compact, by Lemma 3, v_n is lower semicontinuous for each $n = 0, 1, \ldots$. Furthermore, the sequence $\{v_n\}$ is increasing, so that we formally define $v_\infty(x) := \lim_{n\uparrow\infty} v_n(x)$, which, is thus, also lower semicontinuous.

https://doi.org/10.1239/jap/1421763321 Published online by Cambridge University Press

Let $x \in S$ be arbitrarily fixed. It is easy to see from the monotone convergence theorem that $v_\infty(x) \le (c(x,a)/(\alpha + w(x))) + \int_S v_\infty(y) Q(\mathrm{d}y \mid x, a)$ for each $a \in A(x)$, and thus, $v_\infty(x) \le \inf_{a\in A(x)}\{(c(x,a)/(\alpha + w(x))) + \int_S v_\infty(y) Q(\mathrm{d}y \mid x, a)\}$. For the opposite direction, without loss of generality, we assume that $v_\infty < \infty$. For each fixed $m \le n-1$,

$$\begin{aligned} v_\infty(x) &\ge v_n(x) \\ &= \inf_{a\in A(x)} \left\{ \frac{c(x,a)}{\alpha + w(x)} + \int_S v_{n-1}(y) Q(\mathrm{d}y \mid x, a) \right\} \\ &= \frac{c(x,a_n)}{\alpha + w(x)} + \int_S v_{n-1}(y) Q(\mathrm{d}y \mid x, a_n) \\ &\ge \frac{c(x,a_n)}{\alpha + w(x)} + \int_S v_m(y) Q(\mathrm{d}y \mid x, a_n), \end{aligned} \tag{22}$$

where $a_n \in A(x)$ are the corresponding minimizers, whose existence is ensured by Lemma 3, and the last inequality is due to that $\{v_n\}$ is an increasing sequence. Having noted that $c(x,a)/(\alpha + w(x)) + \int_S v_m(y) Q(\mathrm{d}y \mid x, a)$ is $\mathbb{K}$-inf-compact, and $v_\infty(x) < \infty$, we see that the sequence $\{a_n\}$ admits some limit point $a^* \in A(x)$. Assume without loss of generality that $a_n \to a^*$, otherwise we can take the corresponding subsequence. By passing to the limit as $n \to \infty$ on the both sides of (22) and the lower semicontinuity of the involved functions, we obtain $v_\infty(x) \ge c(x,a^*)/(\alpha + w(x)) + \int_S v_m(y) Q(\mathrm{d}y \mid x, a^*)$. Further passing to the limit as $m \to \infty$ on the both sides of the above inequality yields $v_\infty(x) \ge (c(x,a^*)/(\alpha + w(x))) + \int_S v_\infty(y) Q(\mathrm{d}y \mid x, a^*) \ge \inf_{a\in A(x)}\{(c(x,a)/(\alpha + w(x))) + \int_S v_\infty(y) Q(\mathrm{d}y \mid x, a)\}$. Hence, in combination with the other direction as proved earlier, we see that v_∞ is a nonnegative measurable (in fact, lower semicontinuous) solution to (7). This, by virtue of [2, Proposition 9.16], shows $v_\infty(x) = W_\alpha(x)$, and thus, the lower semicontinuity of W_α follows. Consequently, there exists a deterministic stationary policy φ^* such that

$$\begin{aligned} W_\alpha(x) &= \inf_{a\in A(x)} \left\{ \frac{c(x,a)}{\alpha + w(x)} + \frac{w(x)}{w(x)+\alpha} \int_S W_\alpha(y) \left(\frac{q(\mathrm{d}y \mid x, a)}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right) \right\} \\ &= \frac{c(x,\varphi^*(x))}{\alpha + w(x)} + \frac{w(x)}{w(x)+\alpha} \int_S W_\alpha(y) \left(\frac{q(\mathrm{d}y \mid x, \varphi^*(x))}{w(x)} + \mathbf{1}\{x \in \mathrm{d}y\} \right). \end{aligned}$$

Evidently, this policy satisfies $W_\alpha(x) = W_\alpha(x, \varphi^*)$.

Acknowledgements

The author is thankful to the helpful comments and remarks received from an anonymous referee and an editor.

References

[1] Berberian, S. K. (1999). *Fundamentals of Real Analysis*. Springer, New York.

[2] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control*. Academic Press, New York.

[3] Cavazos-Cadena, R. (1991). A counterexample on the optimality equation in Markov decision chains with the average cost criterion. *Syst. Control Lett.* **16,** 387–392.

[4] Cavazos-Cadena, R. and Salem-Silva, F. (2010). The discunted method and equivalence of average criteria for risk-sensitive Markov decision processes on Borel spaces. *Appl. Math. Optimization* **61,** 167–190.

[5] Costa, O. L. V. and Dufour, F. (2012). Average control of Markov decision processes with Feller transition probabilities and general acton spaces. *J. Math. Anal. Appl.* **396,** 58–69.

[6] 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.
[7] Feinberg, E. A. and Lewis, M. E. (2007). Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem. *Math. Operat. Res.* **32,** 769–783.
[8] Feinberg, E. A., Kasyanov, P. O. and Zadoianchuk, N. V. (2012). Average cost Markov decision processes with weakly continuous transition probabilities. *Math. Operat. Res.* **37,** 591–607.
[9] 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.
[10] Feinberg, E. A., Kasyanov, P. O. and Zadoianchuk, N. V. (2013). Fatou's lemma for weakly converging probabilities. Preprint, Department of Applied Mathematics and Statistics, State University of New York at Stony Brook. Available at http://arxiv.org/abs/1206.4073v2.
[11] 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.
[12] Gíhman, Ĭ. Ī. and Skorohod, A. V. (1975). *The Theory of Stochastic Processes. II*. Springer, New York.
[13] Guo, X. (2007). Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces. *Math. Operat. Res.* **32,** 73–87.
[14] Guo, X. and Hernández-Lerma, O. (2003). Drift and monotonicity conditions for continuous-time controlled Markov chains with an average criterion. *IEEE Trans. Automatic Control* **48,** 236–245.
[15] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes: Theory and Applications*. Springer, Berlin.
[16] Guo, X. and Liu, K. (2001). A note on optimality conditions for continuous-time Markov decision processes with average cost criterion. *IEEE Trans. Automatic Control* **46,** 1984–1989.
[17] Guo, X. and Rieder, U. (2006). Average optimality for continuous-time Markov decision processes in Polish spaces. *Ann. Appl. Prob.* **16,** 730–756.
[18] Guo, X. and Ye, L. (2010). New discount and average optimality conditions for continuous-time Markov decision processes. *Adv. Appl. Prob.* **42,** 953–985.
[19] Guo, X. and Zhang, Y. (2013). Generalized discounted continuous-time Markov decision processes. Preprint. Available at http://arxiv.org/abs/1304.3314.
[20] 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.
[21] 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.
[22] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes*. Springer, New York.
[23] 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.
[24] Jaśkiewicz, A. (2009). Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities. *J. Optimization Theory Appl.* **141,** 321–347.
[25] Jaśkiewicz, A. and Nowak, A. S. (2006). On the optimality equation for average cost Markov control processes with Feller transition probabilities. *J. Math. Anal. Appl.* **316,** 495–509.
[26] Jaśkiewicz, A. and Nowak, A. S. (2006). Optimality in Feller semi-Markov control processes. *Operat. Res. Lett.* **34,** 713–718.
[27] Kitaev, M. Yu. and Rykov, V. V. (1995). *Controlled Queueing Systems*. CRC, Boca Raton, FL.
[28] Kitayev, M. Yu. (1986). Semi-Markov and jump Markov controlled models: average cost criterion. *Theory. Prob. Appl.* **30,** 272–288.
[29] Kuznetsov, S. E. (1981). Any Markov process in a Borel space has a transition function. *Theory. Prob. Appl.* **25,** 384–388.
[30] 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.
[31] Piunovskiy, A. and Zhang, Y. (2012). The transformation method for continuous-time Markov decision processes. *J. Optimization Theory Appl.* **154,** 691–712.
[32] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). *Selected Topics on Continuous-time Controlled Markov Chains and Markov Games*. Imperial College Press, London.
[33] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.
[34] Zhu, Q. (2008). Average optimality for continuous-time Markov decision processes with a policy iteration approach. *J. Math. Anal. Appl.* **339,** 691–704.