Constrained continuous-time Markov decision processes on the finite horizon*

Xianping Guo[a], Yonghui Huang[a], and Yi Zhang[b] †
[a] School of Mathematics and Computational Science,
Sun Yat-Sen University, Guangzhou, 510275, P.R. China

[b] Department of Mathematical Sciences,
University of Liverpool, Liverpool, L69 7ZL, UK

Abstract: This paper studies the constrained (nonhomogeneous) continuous-time Markov decision processes on the finite horizon. The performance criterion to be optimized is the expected total reward on the finite horizon, while N constraints are imposed on similar expected costs. Introducing the appropriate notion of the occupation measures for the concerned optimal control problem, we establish the following under some suitable conditions: (a) the class of Markov policies is sufficient; (b) every extreme point of the space of performance vectors is generated by a deterministic Markov policy; and (c) there exists an optimal Markov policy, which is a mixture of no more than $N + 1$ deterministic Markov policies.

Keywords. Continuous-time Markov decision process, constrained-optimality, finite horizon, mixture of $N + 1$ deterministic Markov policies, occupation measure.

Mathematics Subject Classification. 90C40, 60J27

1 Introduction

This paper considers a nonhomogeneous continuous-time Markov decision process (CTMDP) in a Borel state space on a finite time horizon with N constraints.

To the best of our knowledge, the majority of the current literature on constrained CTMDPs considers an infinite time horizon; see e.g., [5, 9, 10, 13, 14, 25] and [9, 12, 29] dealing with the total discounted and long-run average rewards, respectively. For unconstrained CTMDPs on a finite horizon, we mention e.g., [3, 11, 8, 19, 26, 28], which establish the optimality equation. The constrained optimal problem for a CTMDP on a finite horizon has received less attention, see e.g., [20], which is closely related to the present paper. In [20] for rewards in special forms (linear

*Research supported by NSFC. Y.Zhang's work was carried out with a financial grant from the Research Fund for Coal and Steel of the European Commission, within the INDUSE-2-SAFETY project (Grant No. RFSR-CT-2014-00025).

†X. Guo's email: mcsgxp@mail.sysu.edu.cn; Y. Huang's email (Corresponding author): hyongh5@mail.sysu.edu.cn ; Y. Zhang's email: yi.zhang@liv.ac.uk

in the system state) on a finite state space, the authors reduced the constrained finite horizon CTMDP problem to a deterministic optimal control problem through the Kolmogorov equations, by investigating which, the maximum principle for the CTMDP problem was established. The present paper follows a different method from [20]; our investigations are based on the study of occupation measures, and the reduction of the CTMDP problem to a constrained optimality problem over the set of all occupation measures. Compared to [20], we do not require any special form on the rewards/costs, and the model is in a general Borel state space. Furthermore, our main result asserting the optimality of a Markov policy, which is a mixture of deterministic Markov policies, was not obtained or mentioned in [20].

More precisely, we will deal with the constrained CTMDP on a finite horizon under suitable conditions similar to those imposed in [9, 11, 12, 14, 25, 27], which, except for [11], all deal with CTMDP problems on an infinite time horizon. In particular, our model admits the following: (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 space and the action space are both general Borel spaces; and (4) the performance criterion to be optimized is the expected finite horizon rewards, while N constraints are imposed on similar expected finite horizon costs.

The main results and contributions of the present paper are as follows. First, we introduce the appropriate notion of an occupation measure of a policy for the finite horizon CTMDP. The occupation measure in the present paper is necessarily different from and more detailed than the occupation measure for infinite horizon models; see [13, 14, 25]. The space of occupation measures is characterized, and its convexity and compactness with respect to some appropriate topology are shown under the imposed conditions. The characterization result allows one to rewrite the constrained CTMDP optimal control problem as a constrained static optimization problem over the set of occupation measures. We show that the occupation measure of each given policy coincides with the one of some Markov policy; see Theorem 4.1. Then the compactness result leads to the existence of an optimal policy for the original constrained CTMDP optimal control problem (see Theorem 5.1). Second, we show that each extreme point of the performance vector space is generated by a deterministic Markov policy, and in turn establish the existence of an optimal Markov policy, which is a mixture of no more than $N + 1$ deterministic Markov policies; see Theorem 5.2. Similar results were known in [1, 7, 23] for discrete-time Markov decision processes and [14, 15, 24, 25] for CTMDPs on an infinite time horizon. However, to the best of our knowledge, such results on the optimality of the mixture of deterministic Markov policies have not been reported in the current literature on finite horizon CTMDPs as considered in this paper.

The rest of the paper is organized as follows. In Section 2 we introduce the constrained optimal control problem for the finite horizon CTMDP. After giving some preliminaries in Section 3, the properties of occupation measures are examined in Section 4. The main optimality results on

the existence of a constrained-optimal policy are given in Section 5. We finish the paper with a conclusion in Section 6.

2 Optimal control problem

In what follows, for each Borel space X, we denote its Borel σ-algebra by $\mathcal{B}(X)$. Unless stated otherwise, by measurability we mean the Borel measurability. For each subset Γ of X, denote by Γ^c its complement, and by I_Γ the indicator function. For a finite signed measure μ, $|\mu|$ denotes its total variation.

The nonhomogeneous CTMDP model with N constraints is a collection

$$\mathbb{M} := \{S, A, A(t,x)(t \geq 0, x \in S), q(\cdot|t,x,a), (r_k(t,x,a), g_k(x))_{k=0}^N\}, \tag{2.1}$$

consisting of the following elements:

(a) a nonempty Borel space S equipped with the Borel σ-algebra $\mathcal{B}(S)$, called the state space, whose elements are referred to as the states of a system;

(b) a nonempty Borel space A equipped with the Borel σ-algebra $\mathcal{B}(A)$, called the action space, whose elements are referred to as the actions (or decisions) of a decision-maker (or controller);

(c) a family $\{A(t,x), t \geq 0, x \in S\}$ of nonempty subsets $A(t,x)$ of A, where each $A(t,x)$ denotes the set of actions available to a controller when the system is in state $x \in S$ at time t, and it is assumed that $A(t,x) \in \mathcal{B}(A)$, and there is a measurable mapping $f : [0,\infty) \times S \to A$ such that $f(t,x) \in A(t,x)$ for all $t \geq 0$ and $x \in S$;

(d) transition rates $q(\cdot|t,x,a)$, a Borel measurable signed kernel on S given $[0,\infty) \times \mathbb{K}$, satisfying $0 \leq q(D|t,x,a) < +\infty$ for all $(t,x,a) \in \mathbb{K}$ and $x \notin D \in \mathcal{B}(S)$, being conservative in the sense of $q(S|t,x,a) \equiv 0$ and stable in the sense of

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

where

$$\mathbb{K} := \{(t,x,a) : t \geq 0, x \in S, a \in A(t,x)\}$$

is assumed to be a Borel measurable subset of $[0,\infty) \times S \times A$, and

$$q(t,x,a) := -q(\{x\}|t,x,a) \geq 0$$

for all $(t,x,a) \in \mathbb{K}$. For the future reference, let

$$\tilde{q}(\Gamma|t,x,a) := q(\Gamma - \{x\}|t,x,a)$$

for each $\Gamma \in \mathcal{B}(S)$.

(e) the reward rate r_0 and the cost rates r_k are Borel measurable (real-valued) functions on $\mathbb{K}$, while the Borel measurable functions g_0 and g_k on S, $k = 1, \ldots, N$, denote the terminal reward and cost rates, respectively.

Next, we give an informal description of the evolution of a CTMDP with the model (2.1). Roughly speaking, the controller observes the system state continuously in time. If the system remains at the state x at time t, he/she chooses an action $a \in A(t,x)$ according to some given policy, as a consequence of which, the following happens:

(i) Immediate rewards/costs $r_k(t,x,a)dt$ are received.

(ii) A transition from state x to some state in D (with $x \notin D$) occurs with probability $q(D|t,x,a)dt + o(dt)$, or the system remains at state x with probability $1 - q(x|t,x,a)dt + o(dt)$.

To formalize what is described above, we now describe the construction of a CTMDP. Let $S_\Delta := S \bigcup \{\Delta\}$ (with some isolated point $\Delta \notin S$), $\Omega^0 := (S \times (0,\infty))^\infty$, and the sample space be

$$\Omega := \Omega^0 \bigcup \{(x_0, \theta_1, x_1, \ldots, \theta_k, x_k, \infty, \Delta, \infty, \ldots) |\ x_0 \in S, x_l \in S, \theta_l \in (0,\infty),\ \forall\ 1 \le l \le k, k \ge 1\},$$

and let $\mathcal{F}$ be the Borel σ-algebra on Ω. Then we obtain the measurable space $(\Omega, \mathcal{F})$. For each $k \ge 0$ and $e := (x_0, \theta_1, x_1, \ldots, \theta_k, x_k, \ldots) \in \Omega$, let $h_k(e) := (x_0, \theta_1, x_1, \ldots, \theta_k, x_k)$ denote the k-component internal history, and define

$$T_0(e) := 0, \quad T_{k+1}(e) := \theta_1 + \theta_2 + \ldots + \theta_{k+1}, \quad X_k(e) := x_k.$$

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

$$\xi_t := \begin{cases} x_k, & \text{if } T_k \le t < T_{k+1}, \\ \Delta, & \text{if } t \ge T_\infty, \end{cases}$$

for each $t \ge 0$.

Evidently, T_k ($k \ge 1$) denotes the k-th jump moment of $\{\xi_t\}$, X_{k-1} is the state of the process on $[T_{k-1}, T_k)$, and θ_k plays the role of the sojourn time at state X_{k-1}. We formally put $q(\cdot|t,\Delta,a_\Delta) :\equiv 0$, $r_k(t,\Delta,a_\Delta) :\equiv 0$, $A(t,\Delta) :\equiv \{a_\Delta\}$, $A_\Delta := A \cup \{a_\Delta\}$, where $a_\Delta \notin A$ is an isolated point.

Take the right-continuous family of σ-algebras $\{\mathcal{F}_t\}_{t\ge 0}$ as the internal history of the marked point process $\{T_k, X_k, k \ge 0\}$, that is, $\mathcal{F}_t := \sigma(T_m \le s, X_m \in \Gamma, \Gamma \in \mathcal{B}(S), s \le t, m \ge 0)$. Let $\mathcal{P}$ be the σ-algebra of predictable sets on $\Omega \times [0,\infty)$ related to $\{\mathcal{F}_t\}_{t\ge 0}$, that is, $\mathcal{P} := \sigma(\{\Gamma \times \{0\}, \Gamma \in \mathcal{F}_0\} \cup \{\Gamma \times (s,\infty), \Gamma \in \mathcal{F}_{s-}, s > 0\})$, where $\mathcal{F}_{s-} := \vee_{t<s}\mathcal{F}_t$; see Chapter 4 in [21] for details. A real-valued function on $\Omega \times [0,\infty)$ is called predictable if it is measurable with respect to $\mathcal{P}$.

Definition 2.1. A policy is a $\mathcal{P}$-measurable transition probability $\pi(da|e,t)$ on $\mathcal{B}(A_\Delta)$ from $\Omega \times [0,\infty)$, which is concentrated on $A(t,\xi_{t-})$, where $\xi_{t-} = \lim_{s\uparrow t}\xi_s$. A policy $\pi(da|e,t)$ is called Markov

if there is a stochastic kernel ϕ on A given $[0,\infty)\times S$ such that $\pi(da|e,t)=\phi(da|t,\xi_{t-}(e))$ and $\phi(A(t,x)|t,x)\equiv 1$. We will denote by $\phi=\phi(da|t,x)$ a Markov policy. A Markov policy ϕ is called deterministic Markov whenever there exists a A-valued Borel measurable function $f(t,x)$ on $[0,\infty)\times S$ such that $\phi(da|t,x)$ is a Dirac measure concentrated at $f(t,x)$. Such a deterministic Markov policy will be denoted by f for simplicity.

We denote by Π the set of all policies, by Π_m^r the set of all Markov policies, and by Π_m^d the set of all deterministic Markov policies.

Theorems 4.13 and 4.19 or (4.38) in [21] imply that each policy $\pi(da|e,t)$ can be characterized by the following expression

$$\begin{aligned}\pi(da|e,t) &= I_{\{t=0\}}\pi^0(da|x_0,0)+\sum_{k\geq 0} I_{\{T_k<t\leq T_{k+1}\}}\pi^k(da|x_0,\theta_1,x_1,\ldots,\theta_k,x_k,t-T_k)\\ &\quad +I_{\{t\geq T_\infty\}}\delta_{a_\Delta}(da),\end{aligned} \tag{2.3}$$

where $\pi^0(da|x_0,0)$ is a stochastic kernel on A given S concentrated on $A(0,x_0)$, $\pi^k(k\geq 1)$ are stochastic kernels on A given $(S\times(0,\infty))^{k+1}$ concentrated on $A(t,x_k)$, and $\delta_{a_\Delta}(da)$ denotes the Dirac measure at the point a_Δ.

Evidently, for any policy $\pi\in\Pi$ and $D\in\mathcal{B}(S)$, the random measure

$$m^\pi(D|e,t)dt:=\int_A q(D|t,\xi_{t-},a)\pi(da|e,t)I_{\{\xi_{t-}\notin D\}}dt \tag{2.4}$$

is predictable. Note that $m^\pi(D|e,t)$ in (2.4) defines the jump intensity of the process $\{\xi_t\}$, which together with (2.3) gives the following representation

$$m^\pi(D|e,t)=I_{\{t=0\}}m_0^\pi(D|x_0,0)+\sum_{k\geq 0} I_{\{T_k<t\leq T_{k+1}\}}m_k^\pi(D|x_0,\theta_1,x_1,\ldots,\theta_k,x_k,t-T_k),$$

where $m_k^\pi(D|x_0,\theta_1,x_1,\ldots,\theta_k,x_k,t-T_k):=\int_A q(D|t,x_k,a)\pi^k(da|x_0,\theta_1,\ldots,\theta_k,x_k,t-T_k)I_{\{x_k\notin D\}}$ for $T_k<t\leq T_{k+1}$, $m_0^\pi(D|x_0,0):=\int_A q(D|0,x_0,a)\pi^0(da|x_0,0)I_{\{x_0\notin D\}}$, see [22] for details.

Let the policy π be fixed. By a theorem of Jacod's, given the initial distribution γ on $\mathcal{B}(S)$, there is a unique probability measure P_γ^π on $(\Omega,\mathcal{F})$ under which the random measure m^π defined in the above is the unique dual predictable projection of the random measure on $\mathcal{B}((0,\infty)\times S)$ defined by $\sum_{n\geq 1}\delta_{(T_n,X_n)}(dt,dx)$; see more details and the relevant definitions in Chapter 4 of [21] or [22]. This fact is useful in the proof of Lemma 3.3 below. Let us recall the more explicit construction of the measure P_γ^π on the measurable space $(\Omega,\mathcal{F})$ given in [12, 14, 25]. Let $H_0=S$ and $H_k=S\times((0,\infty]\times S_\Delta)^k, k=1,2,\ldots$. The measure P_γ^π on $H_0=S$ is given by $P_\gamma^\pi(D)=\gamma(D)$ for all $D\in\mathcal{B}(S)$. Suppose that the measure P_γ^π on H_k has been constructed. Actually, P_γ^π will be a measure on $(\Omega,\mathcal{F})$, but here, with slight abuse of notation we use it also to denote its marginal projection onto the space of k-component histories H_k. Then P_γ^π on H_{k+1} is determined by the

following formula:

$$P_\gamma^\pi(\Gamma \times (dt, dx)) := \int_\Gamma P_\gamma^\pi(dh_k) I_{\{\theta_{k+1}<\infty\}} m_k^\pi(dx|h_k, t) e^{-\int_0^t m_k^\pi(S|h_k,v)dv} dt, \tag{2.5}$$
$$P_\gamma^\pi(\Gamma \times (\infty, \Delta)) := \int_\Gamma P_\gamma^\pi(dh_k)\{I_{\{\theta_{k+1}=\infty\}} + I_{\{\theta_{k+1}<\infty\}} e^{-\int_0^\infty m_k^\pi(S|h_k,v)dv}\},$$

where $\Gamma \in \mathcal{B}(H_k)$. According to the Ionescu Tulcea theorem, there exists a unique probability measure P_γ^π on $(\Omega, \mathcal{F})$, which has a projection on H_k satisfying (2.5). Let $\mathbb{E}_\gamma^\pi$ be its corresponding expectation operator.

Let $T \in (0, \infty)$ be a fixed finite terminal time. For each policy $\pi \in \Pi$, we define

$$V(\pi, r_k, g_k) = \mathbb{E}_\gamma^\pi \left[\int_0^T \int_A r_k(t, \xi_t, a)\pi(da|e, t)dt + g_k(\xi_T) \right], \qquad k = 0, 1, \ldots, N, \tag{2.6}$$

provided that the expectations are well defined.

Let the numbers, d_k, $k = 1, 2, \ldots, N$, be the constrained constants. We denote by

$$U = \{\pi \in \Pi : \ V(\pi, r_k, g_k) \le d_k, \quad \text{for } k = 1, \ldots, N\} \tag{2.7}$$

the set of policies satisfying the N constraints. A policy $\pi \in \Pi$ is called feasible if it is in U. Throughout this article, to avoid trivial cases, we suppose that $U \neq \emptyset$, and this assumption will not be mentioned explicitly below. Then, the constrained optimal control problem under consideration is as follows:

$$\text{Maximize} \ \ V(\pi, r_0, g_0) \quad \text{over all} \ \ \pi \in U. \tag{2.8}$$

Definition 2.2. A policy $\pi^* \in U$ is called optimal if

$$V(\pi^*, r_0, g_0) = \sup_{\pi \in U} V(\pi, r_0, g_0). \tag{2.9}$$

The main objective of this paper is to show the existence of a Markov optimal policy, which is a mixture of no more than $N + 1$ deterministic Markov policies; see Section 5.

3 Preliminaries

In this section, we present some assumptions and preliminary facts that are used to prove our main results in the subsequent sections.

Assumption 3.1. There exist a continuous function $\omega \ge 1$ on S and constants $c > 0$, $b \ge 0$, $M > 0$ such that

(i) $\int_S q(dy|t, x, a)\omega(y) \le c\omega(x) + b$, for all $(t, x, a) \in \mathbb{K}$;

(ii) $q^*(x) \le M\omega(x)$ for all $x \in S$, with $q^*(x)$ as in (2.2);

(iii) $|r_k(t,x,a)| \le M\omega(x)$, $|g_k(x)| \le M\omega(x)$ for each $(t,x,a) \in \mathbb{K}$ and $0 \le k \le N$.

(iv) $L := \int_S \omega(x)\gamma(dx) < \infty$, where γ is the given initial distribution.

The above condition guarantees that $T_\infty := \lim_{k\to\infty} T_k$ is infinite almost surely with respect to P_γ^π under each $\pi \in \Pi$, see Lemma 3.1. This fact is essential to the validity of the representation (3.5) in the proof of Lemma 3.3 below. Parts (iii) and (iv), together with Assumption 3.2, imply that the Dynkin formula is applicable to the class of functions of interest, see also Remark 3.1 below.

The following two lemmas summarize some consequences of Assumption 3.1.

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

(a) $\mathbb{E}_\gamma^\pi[\omega(\xi_t)] \le e^{ct}[L + \frac{b}{c}]$, for each $t \ge 0$, with L as in Assumption 3.1(iv);

(b) $P_\gamma^\pi(\xi_t \in D) = \gamma(D) + \mathbb{E}_\gamma^\pi\big[\int_0^t \int_A q(D|s,\xi_{s-},a)\pi(da|e,s)ds\big]$, for each $t \ge 0$ and $D \in \mathcal{B}(S)$;

(c) $P_\gamma^\pi(\xi_t \in S) = 1$, for each $t \ge 0$.

Proof. It follows from Lemma 3.1 in [11], see also Proposition 3.1 in [25]. □

Lemma 3.2. *Suppose that Assumption 3.1 holds. Then, for each $k = 0, 1, \ldots, N$,*

$$|V(\pi, r_k, g_k)| \le (T+1)Me^{cT}[L + \frac{b}{c}] \quad \forall\ \pi \in \Pi.$$

Proof. For each $\pi \in \Pi$ and $0 \le k \le N$, by Lemma 3.1(a) and Assumption 3.1(iii)

$$\begin{aligned} |V(\pi, r_k, g_k)| &= \big|\mathbb{E}_\gamma^\pi\big[\int_0^T \int_A r_k(t,\xi_t,a)\pi(da|e,t)dt + g_k(\xi_T)\big]\big| \\ &\le \int_0^T M\mathbb{E}_\gamma^\pi[\omega(\xi_t)]dt + M\mathbb{E}_\gamma^\pi[\omega(\xi_T)] \\ &\le M\int_0^T e^{ct}\big[L + \frac{b}{c}\big]dt + M\big[e^{cT}L + \frac{b}{c}e^{cT}\big] \\ &\le (T+1)Me^{cT}[L + \frac{b}{c}]. \end{aligned}$$

□

We introduce some additional conditions important for the validity of the relevant statement in Lemma 3.3 below.

Assumption 3.2. Let the function ω be as in Assumption 3.1. There exist a continuous function $\omega' \ge 1$ on S and constants $c' > 0$, $b' \ge 0$ and $M' > 0$ such that

(i) $\int_S \omega'(y)q(dy|t,x,a) \le c'\omega'(x) + b'$, for all $(t,x,a) \in \mathbb{K}$;

(ii) $\omega(x)(1+q^*(x)) \le M'\omega'(x)$, with $q^*(x)$ as in (2.2);

(iii) $L' := \int_S \omega'(x)\gamma(dx) < \infty$.

Remark 3.1. *The role of Assumption 3.2 is for the finiteness of* $\mathbb{E}^\pi_\gamma[\omega(\xi_t)q^*(\xi_t)]$ *for* $t \ge 0$*; see the assertions in (3.2) and (3.3) in proving Lemma 3.3 below. However, when the transition rates or the reward functions are bounded (i.e.,* $\sup_{(t,x,a)\in\mathbb{K}} |r_k(t,x,a)| < \infty$, $\sup_{x\in S} |g_k(x)| < \infty$*), Assumption 3.2 is not required at least for the relevant statement in Lemma 3.3 below.*

Let $I := [0,T]$. Given any function $\bar\omega \ge 1$ on S, a function φ on $I \times S$ is called $\bar\omega$-bounded if the $\bar\omega$-weighted norm of φ, $\|\varphi\|_{\bar\omega} := \sup_{(t,x)\in I\times S} \frac{|\varphi(t,x)|}{\bar\omega(x)}$, is finite. We denote by $B_{\bar\omega}(I\times S)$ the Banach space of all $\bar\omega$-bounded Borel measurable functions on $I\times S$, and by $C_b(I\times S)$ the space of all bounded continuous functions on $I\times S$. Obviously, $C_b(I\times S) \subset B_1(I\times S)$. Let

$$K := \{(t,x,a) : t \in [0,T],\ x \in S,\ a \in A(t,x)\}.$$

Since $K = \mathbb{K}\bigcap([0,T]\times S\times A$, and $\mathbb{K} \in \mathcal{B}([0,\infty)\times S\times A)$ by the assumption above, K is also a Borel measurable subset of $[0,\infty)\times S\times A)$. The class of $\bar\omega$-bounded Borel measurable functions on K, denoted by $B_{\bar\omega}(K)$, is similarly defined.

Consider a function $\varphi \in B_\omega(I\times S)$. We mention that if $\varphi(t,x)$ is absolutely continuous in $t\in I$ for each $x\in S$, then there is some measurable function φ' on $I\times S$ satisfying

$$\varphi(s,x) - \varphi(u,x) = \int_u^s \varphi'(t,x)dt,\ \forall\ x\in S,\ 0\le s\le u\le T.$$

Then for each $x\in S$, the function $\varphi'(t,x)$ on $I\times S$ coincides with the partial derivative of the function $\varphi(t,x)$ in $t\in I$ apart from on a null set $L_\varphi(x)\subset I$ with respect to the Lebesgue measure. With ω and ω' as in Assumption 3.2, let $C^{1,0}_{\omega,\omega'}(I\times S) := \{\varphi\in B_\omega(I\times S)$: for each $x\in S$, $\varphi(t,x)$ is absolutely continuous in $t\in I$, and $\varphi' \in B_{\omega+\omega'}(I\times S)\}$.

Lemma 3.3. Suppose Assumptions 3.1(i, ii) and 3.2 are satisfied. Then, the following assertions hold.

(a) (Dynkin's formula): for each $\varphi \in C^{1,0}_{\omega,\omega'}(I\times S)$, under every $\pi\in\Pi$,

$$\mathbb{E}^\pi_\gamma\left[\int_0^T \left(\varphi'(t,\xi_t) + \int_S\int_A \varphi(t,x)q(dx|t,\xi_t,a)\pi(da|e,t)\right)dt\right] = \mathbb{E}^\pi_\gamma[\varphi(T,\xi_T)] - \varphi(0,\gamma),$$

where $\varphi(0,\gamma) := \int_S \varphi(0,x)\gamma(dx)$.

(b) For each $\pi\in\Pi$ and $h\in B_w(I\times S)$,

$$\mathbb{E}^\pi_\gamma\left[\int_0^T\int_S\int_t^T\int_A h(s,x)q(dx|t,\xi_t,a)\pi(da|e,t)dsdt\right] = \mathbb{E}^\pi_\gamma\left[\int_0^T h(t,\xi_t)dt\right] - \int_0^T h(t,\gamma)dt,$$

where $h(t,\gamma) := \int_S h(t,x)\gamma(dx)$ for all $t\ge 0$.

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

$$|\varphi(s,y)| \leq \|\varphi\|_\omega \omega(y), \text{ and } |\varphi'(s,y)| \leq \|\varphi'\|_{\omega+\omega'}(\omega(y)+\omega'(y)) \tag{3.1}$$

for all $s \in I$ and $y \in S$. Under the conditions of the statement, we have

$$\begin{aligned}
\int_A \int_S |q|(dx|s,y,a)|\varphi(s,x)|\pi(da|e,s) &\leq \|\varphi\|_\omega \Big[\int_{S-\{y\}} \int_A \omega(x) q(dx|s,y,a)\pi(da|e,s) + \omega(y) q^*(y) \Big] \\
&\leq \|\varphi\|_\omega \Big[\int_S \int_A \omega(x) q(dx|s,y,a)\pi(da|e,s) + 2\omega(y) q^*(y) \Big] \\
&\leq \|\varphi\|_\omega \big[c\omega(y) + 2M'\omega'(y) + b \big] \quad \forall\ (s,y) \in I \times S. \qquad (3.2)
\end{aligned}$$

Thus, (3.2) and Lemma 3.1(a) give

$$\begin{aligned}
&\int_0^T \mathbb{E}^\pi_\gamma \left[\int_A \int_S |q|(dx|s,\xi_s,a)|\varphi(s,x)|\pi(da|e,s) \right] ds \\
&\leq \|\varphi\|_\omega \int_0^T \mathbb{E}^\pi_\gamma \Big[c\omega(\xi_s) + b + 2M'\omega'(\xi_s) \Big] ds \\
&\leq T\|\varphi\|_\omega [(c+b)e^{cT} L + b + 2M' e^{c'T}(L' + \frac{b'}{c'})] < \infty. \qquad (3.3)
\end{aligned}$$

Moreover, by (3.1) we have

$$\int_0^T |\varphi'(s,\xi_s)| ds \leq \|\varphi'\|_{\omega+\omega'} \int_0^T (\omega(\xi_s) + \omega'(\xi_s)) ds,$$

which, together with Lemma 3.1(a) (with ω being replaced by $(\omega+\omega')$ here), gives

$$\mathbb{E}^\pi_\gamma \left[\int_0^T |\varphi'(s,\xi_s)| ds \right] \leq \|\varphi'\|_{\omega+\omega'} T e^{(c+c')T} [L + L' + \frac{b+b'}{c+c'}] < \infty. \tag{3.4}$$

Since the process is nonexplosive, similarly to [2] with a deterministic setup, we write that (almost surely with respect to P^π_γ) for each $0 \leq t \leq T$,

$$\varphi(t,\xi_t) = \varphi(0,x) + \int_0^t \varphi'(s,\xi_s) ds + \sum_{n \geq 1} \int_{(0,t]} \Delta\varphi(s,\xi_s) \delta_{T_n}(ds) \tag{3.5}$$

with $\Delta\varphi(s,\xi_s) := \varphi(s,\xi_s) - \varphi(s,\xi_{s-})$. Then, because the random measure m^π is the dual predictable projection of the random measure $\sum_{n\geq 1} \delta_{(T_n,X_n)}(dt,dx)$ on $\mathcal{B}((0,\infty)\times S)$ under P^π_γ, we take expectation in both sides of the above equality and obtain that

$$\begin{aligned}
&\mathbb{E}^\pi_\gamma [\varphi(T,\xi_T)] \\
= \ & \varphi(0,\gamma) + \mathbb{E}^\pi_\gamma \left[\int_0^T \varphi'(s,\xi_s) ds \right] + \mathbb{E}^\pi_\gamma \left[\sum_{n\geq 1} \int_{(0,T]} \Delta\varphi(s,\xi_s) \delta_{T_n}(ds) \right] \\
= \ & \varphi(0,\gamma) + \mathbb{E}^\pi_\gamma \left[\int_0^T \varphi'(s,\xi_s) ds \right] + \mathbb{E}^\pi_\gamma \left[\sum_{n\geq 1} \int_S \int_{(0,T]} (\varphi(s,y) - \varphi(s,\xi_{s-})) \delta_{(T_n,X_n)}(ds,dy) \right]
\end{aligned}$$

$$= \varphi(0,\gamma) + \mathbb{E}^{\pi}_{\gamma}\left[\int_0^T \varphi'(s,\xi_s)ds\right] + \mathbb{E}^{\pi}_{\gamma}\left[\int_S\int_{(0,T]}(\varphi(s,y)-\varphi(s,\xi_{s-}))m^{\pi}(dy|e,s)ds\right]$$
$$= \varphi(0,\gamma) + \mathbb{E}^{\pi}_{\gamma}\left[\int_0^T \varphi'(s,\xi_s)ds\right] + \mathbb{E}^{\pi}_{\gamma}\left[\int_S\int_{(0,T]}\int_A \varphi(s,y)q(dy|s,\xi_{s-},a)\pi(da|e,s)ds\right].$$

Here, integrability results such as (3.3) and (3.4) validate all the involved operations. Moreover, for every $e \in \Omega$, $\xi_{s-}(e) = \xi_s(e)$ on $(0,T]$ except finite time points. Hence, part (a) follows.

(b) For each $(t,x) \in I \times S$ and $h \in B_w(I \times S)$, let $\varphi(t,x) := \int_t^T h(s,x)ds$. Then, we have

$$\varphi \in C^{1,0}_{\omega,\omega'}(I \times S),\ \varphi'(t,x) = -h(t,x),\ \varphi(0,x) = \int_0^T h(s,x)ds, \text{and } \varphi(T,x) = 0,$$

which, together with (a), implies (b). □

Under $\phi \in \Pi^r_m$, $\{\xi_t, t \geq 0\}$ is a pure jump Markov process with respect to the probability space $(\Omega, \mathcal{F}, P^{\phi}_{\gamma})$. We denote by $p^{\phi}(t,x;s,D)$ the Feller's transition function of $\{\xi_t, t \geq 0\}$, which satisfies

$$p^{\phi}(t,x;s,D) = P^{\phi}_{\gamma}(\xi_s \in D|\xi_t = x), \forall\ x \in S,\ D \in \mathcal{B}(S),\ s \geq t \geq 0,$$

see [6]. For each $x \in S$, $t \in [0,T]$ and $\phi = \phi(da|t,x) \in \Pi^r_m$, we introduce that for each measurable function h on $\mathbb{K}$,

$$h(s,x,\phi) := \int_A h(s,x,a)\phi(da|s,x)$$

provided that the right hand side is well defined, and put

$$V(\phi, r_k, 0; t, x) := \int_S\int_t^T r_k(s,y,\phi)p^{\phi}(t,x;s,dy)ds. \quad k = 0,1,\ldots,N. \tag{3.6}$$

Lemma 3.4. Suppose that Assumptions 3.1 and 3.2(i) hold. For any Markov policy $\phi \in \Pi^r_m$ and $0 \leq k \leq N$, $V(\phi, r_k, 0; t, x)$ is a solution of the following equation

$$\varphi'(t,x) + r_k(t,x,\phi) + \int_S \varphi(t,y)q(dy|t,x,\phi) = 0 \quad \forall\ t \in L^c_{\varphi}(x),\ x \in S, \tag{3.7}$$

with the boundary condition $\varphi(T,x) = 0$ for each $x \in S$. Here $q(D|t,x,\phi) := \int_A q(D|t,x,a)\phi(da|t,x)$ for all $x \in S$, $D \in \mathcal{B}(S)$ and $t \geq 0$.

Proof. By the backward Kolmogorov equation (e.g. Theorem 3.1 in [6]) , we have

$$p^{\phi}(t,x;s,D) = \int_t^s\int_S q(dz|v,x,\phi)p^{\phi}(v,z;s,D)dv + \delta_{\{x\}}(D). \tag{3.8}$$

On the other hand, for each $x \in S$ and $t \geq 0$,

$$\int_t^T\int_S |r_k(s,y,\phi)|\int_t^s\int_S |q|(dz|v,x,\phi)p^{\phi}(v,z;s,dy)dvds$$

$$
\begin{aligned}
&= \int_t^T \int_v^T \int_S |q|(dz|v,x,\phi) \int_S |r_k(s,y,\phi)| p^{\phi}(v,z;s,dy)dsdv \\
&= \int_t^T \int_S |q|(dz|v,x,\phi) V(\phi,|r_k|,0;v,z)dv \\
&\leq \int_0^T \int_S |q|(dz|v,x,\phi)(T+1)Me^{cT}[\omega(z)+\frac{b}{c}]dv \\
&\leq T(T+1)Me^{cT}[c\omega(x)+b+2\omega(x)q^*(x)+\frac{2b}{c}q^*(x)] < \infty.
\end{aligned}
$$

Thus, using Fubini's theorem, by (3.6) and (3.8) we have

$$
\begin{aligned}
V(\phi,r_k,0;t,x) &= \int_t^T \int_S r_k(s,y,\phi) p^{\phi}(t,x;s,dy)ds \\
&= \int_t^T \int_S r_k(s,y,\phi) \int_t^s \int_S q(dz|v,x,\phi) p^{\phi}(v,z;s,dy)dvds + \int_t^T r_k(s,x,\phi)ds \\
&= \int_t^T \int_v^T \int_S q(dz|v,x,\phi) \int_S r_k(s,y,\phi) p^{\phi}(v,z;s,dy)dsdv + \int_t^T r_k(s,x,\phi)ds \\
&= \int_t^T \left[\int_S q(dz|v,x,\phi) V(\phi,r_k,0;v,z) \right] dv + \int_t^T r_k(s,x,\phi)ds,
\end{aligned}
$$

and so (3.7) is verified. □

4 Occupation measures

In this section, we introduce the occupation measure of a policy for the finite horizon CTMDP, and present some basic properties of the space of occupation measures.

Definition 4.1. *For each $\pi \in \Pi$, the occupation measure η^{π} of π on K, is defined by*

$$
\eta^{\pi}(dt,dx,da) := \mathbb{E}_{\gamma}^{\pi}[I_{\{\xi_t \in dx\}} \pi(da|e,t)]dt. \tag{4.1}
$$

Note that $\eta^{\pi}(K) = T$, and so $\{\eta^{\pi}, \pi \in \Pi\}$ is a bounded family of measures on $\mathcal{B}(K)$.

Remark 4.1. *For the sake of comparisons, we mention that the occupation measure for discounted models on an infinite time horizon in [13, 14, 25] takes the form of*

$$
\eta^{\pi}(dx,da) = \alpha \int_0^{\infty} e^{-\alpha t} \mathbb{E}_{\gamma}^{\pi}[I_{\{\xi_t \in dx\}} \pi(da|e,t)]dt
$$

with a constant discount factor α. Evidently, the occupation measure for the finite horizon CTMDP as considered in this paper is more detailed.

Now we can rewrite $V(\pi, r_k, g_k)$ as an integral with respect to the occupation measure of π as follows.

Lemma 4.1. Suppose that Assumptions 3.1 and 3.2 hold. Then, for each $\pi \in \Pi$ and $0 \le k \le N$,

$$V(\pi, r_k, g_k) = \int_K H_k(t,x,a)\eta^\pi(dt,dx,da),$$

where

$$H_k(t,x,a) := r_k(t,x,a) + \int_S g_k(y)q(dy|t,x,a) + \frac{1}{T}\int_S g_k(y)\gamma(dy). \tag{4.2}$$

Proof. It follows from (2.6), (4.1) and Lemma 3.1(b). □

By Lemma 4.1, we can reduce the optimal control problem (2.8) to the following static optimization problem

$$\begin{aligned} &\text{Maximize} \quad \int_K H_0(t,x,a)\eta^\pi(dt,dx,da) \ \text{ over } \pi \in \Pi, \\ &\text{subject to} \quad \int_K H_k(t,x,a)\eta^\pi(dt,dx,da) \le d_k, \ k = 1, \cdots, N. \end{aligned} \tag{4.3}$$

In what follows, let $P(K)$ be the collection of Borel measures η on K such that $\eta(K) = T$. For each $\eta \in P(K)$, let $\bar{\eta}(dt,dx)$ be the marginal of η on $I \times S$, and $\underline{\eta}(dx)$ be the marginal of η on S. Remember, $I = [0,T]$. Lemma 9.4.4 in [17] guarantees the existence of $\phi \in \Pi_m^r$ satisfying

$$\eta(dt,dx,da) =: \bar{\eta}(dt,dx)\phi(da|t,x)$$

on $\mathcal{B}(K)$. We define the following sets

$$\mathcal{D} := \{\eta^\pi : \ \pi \in \Pi\}, \tag{4.4}$$

$$P_{\bar{\omega}}(K) := \left\{\eta \in P(K) : \quad \int_S \bar{\omega}(x)\underline{\eta}(dx) < \infty\right\}, \tag{4.5}$$

where $\bar{\omega} \ge 1$ is a real-valued function on S.

The theorem below characterizes the space of occupation measures.

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

(a) For each $\eta \in P_\omega(K)$, it holds that $\eta \in \mathcal{D}$ if and only if

$$\begin{aligned} &\int_K \left(\int_S \int_t^T q(dy|t,x,a)h(s,y)ds\right)\eta(dt,dx,da) \\ &= \int_0^T \int_S h(s,y)\bar{\eta}(ds,dy) - \int_0^T h(s,\gamma)ds \quad \forall\ h \in C_b(I \times S), \end{aligned} \tag{4.6}$$

i.e.,

$$\bar{\eta}(ds,dy) = \gamma(dy)ds + \int_K I_{[t,T]}(s)q(dy|t,x,a)\eta(dt,dx,da)ds, \tag{4.7}$$

on $\mathcal{B}(I \times S)$.

(b) For each $\pi \in \Pi$, there exists a Markov policy ϕ such that $\eta^\pi = \eta^\phi$.

(c) $\mathcal{D}$ is convex.

Proof. (a) Fix some $\eta \in \mathcal{D}$. Then for some policy $\pi \in \Pi$ it holds that $\eta = \eta^\pi$. Moreover, it follows from Lemma 3.1(a) that $\eta^\pi \in P_\omega(K)$. Thus, for any $h \in C_b(I \times S)$, by the definition of η^π and Lemma 3.3(b) we have

$$\begin{aligned}
&\int_K \left(\int_S \int_t^T q(dy|t,x,a) h(s,y) ds \right) \eta(dt,dx,da) \\
&= \mathbb{E}_\gamma^\pi \left[\int_0^T \int_A \int_S \left(\int_t^T h(s,y) ds \right) q(dy|t,\xi_t,a) \pi(da|e,t) dt \right] \\
&= \mathbb{E}_\gamma^\pi \left[\int_0^T h(t,\xi_t) dt \right] - \int_0^T h(t,\gamma) dt \\
&= \int_0^T \int_S h(t,x) \bar{\eta}^\pi(dt,dx) - \int_0^T h(t,\gamma) dt. \qquad (4.8)
\end{aligned}$$

On the other hand, take any $\eta \in P_\omega(K)$ such that (4.6) (or equivalently (4.7)) holds for η. Then there is a Markov policy ϕ satisfying $\eta(dt,dx,da) = \bar{\eta}(dt,dx)\phi(da|t,x)$. We next show $\eta = \eta^\phi$, which is equivalent to the following:

$$\int_K \tilde{h}(t,x,a)\eta(dt,dx,da) = \int_K \tilde{h}(t,x,a)\eta^\phi(dt,dx,da) \quad \forall\ \tilde{h} \in C_b(K). \qquad (4.9)$$

The rest verifies (4.9). Since $\eta(dt,dx,da) = \bar{\eta}(dt,dx)\phi(da|t,x)$, for each $\tilde{h} \in C_b(K)$, we have

$$\begin{aligned}
\int_K \tilde{h}(t,x,a)\eta(dt,dx,da) &= \int_K \tilde{h}(t,x,a)\bar{\eta}(dt,dx)\phi(da|t,x) \\
&= \int_0^T \int_S \tilde{h}(t,x,\phi)\bar{\eta}(dt,dx),
\end{aligned}$$

which, together with Lemma 3.4 and (4.7) as well as (4.1), gives

$$\begin{aligned}
&\int_K \tilde{h}(t,x,a)\eta(dt,dx,da) \\
&= -\int_0^T \int_S \left[\frac{\partial V(\phi,\tilde{h},0;t,x)}{\partial t} + \int_S V(\phi,\tilde{h},0;t,y) q(dy|t,x,\phi) \right] \bar{\eta}(dt,dx) \quad \text{(by Lemma 3.4)} \\
&= -\int_0^T \int_S \frac{\partial V(\phi,\tilde{h},0;t,x)}{\partial t} \bar{\eta}(dt,dx) + \int_0^T \int_S \int_S \left(\int_t^T \frac{\partial V(\phi,\tilde{h},0;s,y)}{\partial s} ds \right) q(dy|t,x,\phi)\bar{\eta}(dt,dx) \\
&= -\int_0^T \frac{\partial V(\phi,\tilde{h},0;t,\gamma)}{\partial t} dt \qquad \text{(by (4.7))} \\
&= \int_S V(\phi,\tilde{h},0;0,x)\gamma(dx) \qquad (4.10) \\
&= \int_K \tilde{h}(t,x,a)\eta^\phi(dt,dx,da). \qquad \text{(by (4.1)).}
\end{aligned}$$

Thus, (4.9) is proved, and so (a) follows.

Parts (b) and (c) follow from part (a) and its proof of (a), see (4.9). □

Remark 4.2. *Recall that the occupation measures in [13, 14, 24, 25] for infinite horizon discounted cases are characterized, under similar conditions as in the present paper, by*

$$\alpha\underline{\eta}(dy) = \alpha\gamma(dy) + \int_S \int_A q(dy|x,a)\eta(dx,da),$$

where homogeneous transition rates $q(\cdot|x,a)$ are treated, c.f. (4.7) in the previous theorem.

Definition 4.2. For each $\bar{\omega} \geq 1$ on S, the $\bar{\omega}$-weak topology on $P_{\bar{\omega}}(K)$ is defined as the weakest topology with respect to which, $\int_K u(t,x,a)\eta(dt,dx,da)$ is continuous in $\eta \in P_{\bar{\omega}}(K)$ for each continuous function u on K such that $\sup_{(t,x,a)\in K} \frac{|u(t,x,a)|}{\bar{\omega}(x)} < \infty$. Convergence in the $\bar{\omega}$-weak topology is signified by $\xrightarrow{\bar{\omega}}$.

Let $\mathcal{P}(K)$ be the collection of all the Borel probability measures on K. For each function $\bar{\omega} \geq 1$ on S, we define two mappings, $T_{\bar{\omega}}$ and $T'_{\bar{\omega}}$, as follows:

$$\begin{aligned} T_{\bar{\omega}}: \quad & P_{\bar{\omega}}(K) \longrightarrow \mathcal{P}(K), \qquad \eta \mapsto T_{\bar{\omega}}(\eta), \text{ where } T_{\bar{\omega}}(\eta) \text{ is given by} \\ & T_{\bar{\omega}}(\eta)(dt,dx,da) := \frac{\bar{\omega}(x)\eta(dt,dx,da)}{\int_S \bar{\omega}(y)\underline{\eta}(dy)}; \end{aligned} \tag{4.11}$$

$$\begin{aligned} T'_{\bar{\omega}}: \quad & \mathcal{P}(K) \longrightarrow P_{\bar{\omega}}(K), \qquad \mu \mapsto T'_{\bar{\omega}}(\mu), \text{ where } T'_{\bar{\omega}}(\mu) \text{ is given by} \\ & T'_{\bar{\omega}}(\mu)(dt,dx,da) := T\frac{\frac{1}{\bar{\omega}(x)}\mu(dt,dx,da)}{\int_S \frac{1}{\bar{\omega}(y)}\underline{\mu}(dy)}. \end{aligned} \tag{4.12}$$

(Since $1 \leq \bar{\omega} < \infty$ on S, we have $0 < \int_S \frac{1}{\bar{\omega}(y)}\bar{\mu}(dy) < \infty$ for any $\mu \in P(K)$, and thus the mappings $T_{\bar{\omega}}$ and $T'_{\bar{\omega}}$ are well defined.)

Lemma 4.2. For each continuous function $\bar{\omega} \geq 1$ on S, $\mathcal{P}(K)$ (endowed with the usual weak topology) and $P_{\bar{\omega}}(K)$ (endowed with the $\bar{\omega}$-weak topology) are homeomorphic with $T_{\bar{\omega}}$ being a homeomorphism.

Proof. See [25]. □

As a consequence, $P_{\bar{\omega}}(K)$ endowed with the $\bar{\omega}$-weak topology is metrizable, provided that the function $\bar{\omega} \geq 1$ is continuous.

Specially, taking $\bar{\omega} = \omega + \omega'$ and ω respectively (with ω and ω' as in Assumption 3.2), we have the following lemma.

Lemma 4.3. Under Assumptions 3.1 and 3.2, if $\int_S f(y)\tilde{q}(dy|t,x,a)$ is continuous in $(t,x,a) \in K$ for each bounded continuous function f on S, then $\mathcal{D}$ is closed in $P_{\omega'}(K)$ and in $P_{\omega}(K)$. Here and below, $P_{\omega'}(K)$ and $P_{\omega}(K)$ are endowed with the ω'-weak topology and the ω-weak topology, respectively.

Proof. Note that $\mathcal{D} \subseteq P_\omega(K)$. We only show that $\mathcal{D}$ is closed in $P_\omega(K)$; the other case is absolutely similar. Take an arbitrary sequence $\{\eta_m\}$ in $\mathcal{D}$ such that $\eta_m \xrightarrow{\omega} \eta_0 \in P_\omega(K)$. Let $\pi_m \in \Pi$ be such that $\eta_m = \eta^{\pi_m}$. Then, under Assumptions 3.1 and 3.2, by Lemma 3.1 we have

$$\int_S \omega(x)\underline{\eta}_m(dx) = \int_0^T \mathbb{E}_\gamma^{\pi_m}[\omega(\xi_t)]dt \le Te^{cT}[L + \frac{b}{c}] =: M^* < \infty \quad \forall\, m \ge 1, \tag{4.13}$$

which, together with $\eta_m \xrightarrow{\omega} \eta_0$, implies

$$\int_S \omega(x)\underline{\eta}_0(dx) = \lim_{m\to\infty} \int_S \omega(x)\underline{\eta}_m(dx) \le M^*. \tag{4.14}$$

Thus, to prove $\eta_0 \in \mathcal{D}$, by Theorem 4.1(a) it suffices to verify (4.6) with η being replaced by η_0. Indeed, for any $h \in C_b(I \times S)$, by $\eta_m \in \mathcal{D}$ and Theorem 4.1(a) we have

$$\begin{aligned} &\int_K \int_S \int_t^T q(dy|t,x,a)h(s,y)ds\eta_m(dt,dx,da) \\ &= \int_S \int_0^T h(t,x)\bar{\eta}_m(dt,dx) - \int_0^T h(t,\gamma)dt \qquad \forall\, m \ge 1. \end{aligned} \tag{4.15}$$

Since $\|h\| := \sup_{(s,x)\in I\times S} |h(s,x)| < \infty$, by Assumption 3.1 we have

$$\int_S \int_t^T |q|(dy|t,x,a)|h(s,y)|ds \le T\|h\|(2q^*(x) + c\omega(x) + b) \le T\|h\|(2M + c + b)\omega(x) \quad \forall\, x \in S.$$

Moreover, it follows from the dominated convergence theorem that $\int_S \int_t^T q(dy|t,x,a)h(s,y)ds$ is continuous in $(t,x,a) \in K$. Thus, letting $m \to \infty$ in (4.15), we see that (4.6) holds for η_0, and so Theorem 4.1(a) implies that $\eta_0 \in \mathcal{D}$. The proof is complete. □

For the compactness of $\mathcal{D}$, as in [12, 14, 25] on the infinite horizon, we further introduce the following condition.

Assumption 4.1. Let ω and ω' be as in Assumption 3.2.

(i) $\int_S f(y)\tilde{q}(dy|t,x,a)$ is continuous in $(t,x,a) \in K$ for each bounded continuous function f on S; and $\int_S g(y)\tilde{q}(dy|t,x,a)$ is continuous in $a \in A(t,x)$ for each $(t,x) \in I \times S$ and bounded measurable function g on S.

(ii) There exists an increasing sequence of compact subsets (K_m) of K satisfying $\bigcup_m K_m = K$ and $\lim_{m\to\infty} \inf_{(t,x,a)\in K\setminus K_m} \frac{\omega'(x)}{\omega(x)} = \infty$, where $\inf \emptyset := \infty$.

Remark 4.3. *Assumption 4.1 implies that $A(t,x)$ is compact for each $(t,x) \in I \times S$; see Lemma 3.10 of [25]. On the other hand, the function $\frac{\omega'(x)}{\omega(x)}$ in Assumption 4.1(ii) is a so-called strictly unbounded or moment function, which plays a role in verifying that $\mathcal{D}$ is sequentially relatively compact; for the details, see the proof of Theorem 4.2 below.*

Theorem 4.2. Suppose that Assumptions 3.1, 3.2 and 4.1 hold. Then, $\mathcal{D}$ is compact in $P_\omega(K)$.

Proof. Since $P_w(K)$ is metrizable and $\mathcal{D}$ is closed (by Lemma 4.3), it suffices to show that $T_w(\mathcal{D})$ is sequentially relatively compact in $\mathcal{P}(K)$ endowed with the usual weak topology. Indeed, for every $\eta^\pi \in \mathcal{D}$, under Assumptions 3.1 and 3.2, by Lemma 4.2 and (4.11), we have

$$\begin{aligned}\int_K \frac{\omega'(x)}{\omega(x)} T_\omega(\eta^\pi)(dt,dx,da) &= \frac{\int_S \omega'(x)\underline{\eta}^\pi(x)}{\int_S \omega(x)\underline{\eta}^\pi(x)} \\ &\le \frac{1}{T}\int_S \omega'(x)\underline{\eta}^\pi(x) \\ &= \frac{1}{T}\int_0^T \mathbb{E}^\pi_\gamma[\omega'(\xi_t)]dt \le e^{c'T}[L' + \frac{b'}{c'}] < \infty. \end{aligned} \tag{4.16}$$

Now (4.16) and the Prokhorov theorem (see Theorem 12.2.15 in [17]) imply that $\{T_\omega(\eta), \eta \in \mathcal{D}\}$ is sequentially relatively compact in $\mathcal{P}(K)$, and so is $\mathcal{D}$ in $P_\omega(K)$ (by Lemma 4.2 with $\bar{\omega} = \omega$). □

5 Existence of optimal policies

This section establishes the existence of a Markov optimal policy, which is a mixture of no more than $N+1$ deterministic Markov policies.

Assumption 5.1. The functions $r_k(t,x,a)$, $g_k(x)$ $(k = 0,1,\ldots,N)$, and $\int_S \omega(y)\tilde{q}(dy|t,x,a)$ are continuous in $(t,x,a) \in K$. Furthermore, one of the following conditions (i) and (ii) holds:

(i) Either q^* or each of the functions g_k is bounded on S.

(ii) There exists a function $\omega'' \ge 1$ on S and constants $c'' > 0$, $b'' \ge 0$ and $M'' \ge 0$ such that

1) $\int_S \omega''(y)q(dy|t,x,a) \le c''\omega''(x) + b''$, $\forall\ (t,x,a) \in \mathbb{K}$;

2) $L'' := \int_S w''(x)\gamma(dx) < \infty$;

3) There exists an increasing sequence of compact subsets (K'_m) of K satisfying

$$\lim_{m\to\infty} \inf_{(t,x,a)\in K\setminus K'_m} \frac{\omega''(x)}{\omega'(x)} = \infty$$

and $\bigcup_m K'_m = K$;

4) $\omega'(x)(1+q^*(x)) \le M''\omega''(x)$ for all $x \in S$,

where ω, ω' are as in Assumption 3.2.

Suppose Assumptions 3.1, 3.2 and 4.1 are satisfied. Then under the first part of Assumption 5.1, one can show that $\int_S f(y)q(dy|t,x,a)$ is continuous in $(t,x,a) \in K$ for each w-bounded continuous function g on S; the reasoning is similar to the one in the proof of Lemma 8.5.5 in [17]. If additionally

Assumption 5.1(i) is satisfied, then H_k is w-bounded and continuous on K, where H_k is defined by (4.2). The function $\int_K H_k(t,x,a)\eta(dt,dx,da)$ is continuous in $\eta \in \mathcal{D} \subseteq P_\omega(K)$ endowed with the ω-topology. By Theorem 4.2, $\mathcal{D}$ is compact in $P_\omega(K)$. Consequently, problem (4.3) has an optimal solution. Now one can apply Theorem 4.1 for the existence of a Markov optimal policy for problem (2.8). If alternatively, Assumption 5.1(ii) is satisfied, then $\int_S g_k(y)q(dy|t,x,a)$ in the definition of H_k is w'-bounded continuous in $(t,x,a) \in K$, so that $\int_K H_k(t,x,a)\eta(dt,dx,da)$ is continuous in $\eta \in \mathcal{D} \subseteq P_{\omega'}(K)$ endowed with the ω'-topology. On the other hand, applying the same reasoning as in the proof of Theorem 4.2, $\mathcal{D}$ is compact in $P_{\omega'}(K)$ under Assumption 5.1(ii). Therefore, we can again conclude the existence of a Markov optimal policy for problem (2.8).

The above discussions amount to the following statement.

Theorem 5.1. Under Assumptions 3.1, 3.2, 4.1, and 5.1, there exists a Markov optimal policy for problem (2.8).

Definition 5.1. A policy $\pi \in \Pi$ is said to be a mixture of $m+1$ deterministic Markov policies $f_l, l = 0,1,2,\ldots,m$, if

$$\eta^\pi(dt,dx,da) = \sum_{l=0}^{m} p_l \eta^{f_l}(dt,dx,da),$$

where $p_l \geq 0$ for all $0 \leq l \leq m$, and $p_0 + \cdots + p_m = 1$.

Under Assumption 3.1, we consider the space of performance vectors for the model (2.1) with the criteria (2.6):

$$\mathcal{U} := \{(V(\pi, r_0, g_0), \ldots, V(\pi, r_N, g_N)) \mid \pi \in \Pi\}. \tag{5.1}$$

In the proof of the main statement below, we shall make use of the next result, whose proof is available in [18]; see also [11] for the proof of its version in the case of a denumerable state space.

Lemma 5.1. Under Assumptions 3.1, 3.2, 4.1 and the first part of Assumption 5.1, the following assertions hold.

(a) There exists a unique φ in $C^{1,0}_{\omega,\omega}(I \times S)$, and a deterministic Markov policy $f^* \in \Pi^d_m$ satisfying the following optimality equation:

$$\begin{aligned}
&\varphi'(t,x) + \sup_{a \in A(t,x)} [r_0(t,x,a) + \int_S \varphi(t,y)q(y|t,x,a)] = 0, \ \forall\, t \in L^c_\varphi(x), \ x \in S;\\
&\varphi'(t,x) + r_0(t,x,f^*(t,x)) + \int_S \varphi(t,y)q(y|t,x,f^*(t,x)) =\\
&\varphi'(t,x) + \sup_{a \in A(t,x)} [r_0(t,x,a) + \int_S \varphi(t,y)q(y|t,x,a)], \ \forall\, t \in I, \ x \in S;\\
&\varphi(T,x) = g_0(x), \ \forall\, x \in S.
\end{aligned} \tag{5.2}$$

(b) The policy f^* and the function φ in (a) satisfy that $V(f^*, r_0, g_0) = \sup_{\pi\in\Pi} V(\pi, r_0, g_0) = \int_S \varphi(0,x)\gamma(dx)$.

We are in position to present the main statement.

Theorem 5.2. Suppose Assumptions 3.1, 3.2, 4.1 and 5.1 are satisfied. Then the following assertions hold.

(a) The space of performance vectors, $\mathcal{U}$, is nonempty, compact and convex.

(b) Each extreme point of $\mathcal{U}$ (there exists at least one), say v^{ex}, is generated by a deterministic Markov policy, say f, i.e., $v^{ex} = (V(f, r_0, g_0), \ldots, V(f, r_N, g_N))$.

(c) There exists an optimal Markov policy, which is a mixture of $(N+1)$ deterministic Markov policies.

Proof. (a) For each $0 \le k \le N$, $u \in B_{\omega'}(K)$ and $\eta \in \mathcal{D}$, let

$$\langle u, \eta\rangle \;:=\; \int_K u(t,x,a)\eta(dt,dx,da). \tag{5.3}$$

Then, by Lemma 4.1 and Theorem 4.1 we have

$$V(\pi, r_k, g_k) = \langle H_k, \eta^\pi\rangle, \qquad \text{for all } \pi \in \Pi, \tag{5.4}$$

and so

$$\mathcal{U} = \{(\langle H_0, \eta\rangle, \ldots, \langle H_N, \eta\rangle) \mid \eta \in \mathcal{D}\}. \tag{5.5}$$

Since the functions H_k are continuous and ω-bounded (resp., ω'-bounded) on K under Assumption 5.1(i) (resp., Assumption 5.1(ii)), $\mathcal{U}$ is nonempty, convex and compact, because so is $\mathcal{D}$. Hence, (a) is true.

(b) By (a) $\mathcal{U}$ admits at least one extreme point, say v^{ex}. Below we prove that any given extreme point v^{ex} of $\mathcal{U}$ is generated by a deterministic Markov policy by induction with respect to the number of constraints N.

Consider the case of $N = 0$ (i.e., $\mathcal{U} = \{\langle H_0, \eta\rangle \mid \eta \in \mathcal{D}\}$). Then, by the convexity and compactness of $\mathcal{D}$ (proved above), $\mathcal{U} \subset \mathbb{R} := (-\infty, \infty)$ is a bounded closed interval, and the two extreme points of $\mathcal{U}$, denoted by v_{min} and v_{max}, corresponding to the two end points of the closed interval, are given by the optimal values of the following two unconstrained finite-horizon CTMDP problems: $\langle H_0, \eta^\pi\rangle = V(\pi, r_0, g_0) \to \max_{\pi\in\Pi}$ and $\langle H_0, \eta^\pi\rangle = V(\pi, r_0, g_0) \to \min_{\pi\in\Pi}$ respectively. Lemma 5.1 gives the existence of deterministic Markov policies f_1 and f_2 satisfying $v_{max} = \sup_{\pi\in\Pi} V(\pi, r_0, g_0) = V(f_1, r_0, g_0)$, and $v_{min} = \inf_{\pi\in\Pi} V(\pi, r_0, g_0) = V(f_2, r_0, g_0)$. Thus, (b) is true for the case of $N = 0$.

Suppose that (b) is true for the case of $N = n-1$. Then, consider the case of $N = n$. For any extreme point $v^{ex} \in \mathcal{U}$ in this case of $N = n$, by (5.5) we can write $v^{ex} = (\langle H_0, \eta^{\pi^{ex}} \rangle, \ldots, \langle H_n, \eta^{\pi^{ex}} \rangle)$, for some $\pi^{ex} \in \Pi_m^r$. Since v^{ex} is an extreme point of $\mathcal{U}$, it is not in the interior of $\mathcal{U} \subset \mathbb{R}^{n+1}$. So by the supporting hyperplane theorem [4], there exists a hyperplane

$$\mathcal{H} := \left\{ (c_0, c_1, \ldots, c_n) \in \mathbb{R}^{n+1} \mid \sum_{k=0}^{n} \lambda_k c_k = \rho^* \right\}, \tag{5.6}$$

where λ_k and ρ^* are fixed constants defining $\mathcal{H}$ such that

$$\sum_{k=0}^{n} \lambda_k \langle H_k, \eta^{\pi^{ex}} \rangle = \rho^* \geq \sum_{k=0}^{n} \lambda_k v_k \quad \text{for all } (v_0, v_1, \ldots, v_n) \in \mathcal{U}.$$

Here it is without loss of generality to put $\lambda_n \neq 0$ for otherwise one just needs to introduce an appropriate relabeling. This, together with (5.4) and (5.3), implies

$$V(\pi^{ex}, \sum_{k=0}^{n} \lambda_k r_k, \sum_{k=0}^{n} \lambda_k g_k) = \langle \sum_{k=0}^{n} \lambda_k H_k, \eta^{\pi^{ex}} \rangle = \rho^* \geq V(\pi, \sum_{k=0}^{n} \lambda_k r_k, \sum_{k=0}^{n} \lambda_k g_k) \quad \forall\ \pi \in \Pi.$$

This means

$$V(\pi^{ex}, \sum_{k=0}^{n} \lambda_k r_k, \sum_{k=0}^{n} \lambda_k g_k) = \sup_{\pi \in \Pi} V(\pi, \sum_{k=0}^{n} \lambda_k r_k, \sum_{k=0}^{n} \lambda_k g_k) = \rho^*. \tag{5.7}$$

Define the set

$$\mathcal{V} := \mathcal{U} \cap \mathcal{H}, \tag{5.8}$$

which is nonempty, convex and compact. Note that the extreme point v^{ex} is also an extreme point of $\mathcal{V}$ since v^{ex} is on $\mathcal{H}$.

Moreover, for any $\phi \in \Pi_m^r, t \in I$, and $x \in S$, let

$$V^{\vec{\lambda}}(t,x) \quad := \quad \sup_{\phi \in \Pi_m^r} V(\phi, H_{\vec{\lambda}}^n, 0; t, x), \tag{5.9}$$

where, $\vec{\lambda} := (\lambda_0, \ldots, \lambda_n)$, and

$$H_{\vec{\lambda}}^n(t,x,a) := \sum_{k=0}^{n} \lambda_k H_k(t,x,a)$$

for each $(t,x,a) \in K$.

Then, by Lemma 5.1, there exists a policy $f_{\vec{\lambda}} \in \Pi_m^d$ such that

$$V(f_{\vec{\lambda}}, H_{\vec{\lambda}}^n, 0; t, x) = V^{\vec{\lambda}}(t,x) \in C_{\omega,\omega'}^{1,0}(I \times S), \tag{5.10}$$

and

$$V^{\vec{\lambda}\,\prime}(t,x) + H_{\vec{\lambda}}^n(t,x,f_{\vec{\lambda}}) + \int_S V^{\vec{\lambda}}(t,y) q(dy|t,x,f_{\vec{\lambda}})$$

$$= \sup_{a\in A(t,x)} \left(V^{\vec{\lambda}\,\prime}(t,x) + H^n_{\vec{\lambda}}(t,x,a) + \int_S V^{\vec{\lambda}}(t,y) q(dy|t,x,a) \right) = 0 \tag{5.11}$$

for all $x \in S$ and $t \in L^c_{f_{\vec{\lambda}}}(x)$.

By (5.4), (5.7), and Theorem 4.1(b) we have

$$V^{\vec{\lambda}}(0,\gamma) = \int_S V^{\vec{\lambda}}(0,x)\gamma(dx) = \sup_{\phi\in\Pi^r_m} \int_S V(\phi, H^n_{\vec{\lambda}}, 0; 0, x)\gamma(dx) = \rho^*. \tag{5.12}$$

For each $x \in S$, and $t \in I$, let

$$\hat{A}(t,x) := \left\{ a \in A(t,x) : V^{\vec{\lambda}}_t(t,x) + H^n_{\vec{\lambda}}(t,x,a) + \int_S V^{\vec{\lambda}}(t,y) q(dy|t,x,a) = 0 \right\} \tag{5.13}$$

whenever the set on the right hand side is nonempty; and for (t,x) at which that set is empty, we put

$$\hat{A}(t,x) := \{ f_{\vec{\lambda}}(t,x) \},$$

where $f_{\vec{\lambda}}$ is the deterministic Markov policy satisfying (5.10). It holds that for each $x \in S$ and $t \in [0,T]$ that $\emptyset \neq \hat{A}(t,x) \subseteq A(t,x)$. In what follows, if necessary, we always extend $\hat{A}(t,x)$ to $[0,\infty)\times S$ by putting $\hat{A}(t,x) = \{f_{\vec{\lambda}}(t,x)\}$ for each $(t,x) \in (T,\infty)\times S$.

For each $(t,x) \in I \times S$, the set $\hat{A}(t,x) \subseteq A(t,x)$ is compact because for any fixed $(t,x) \in I\times S$, the function

$$G(a) := V^{\vec{\lambda}\,\prime}(t,x) + H^n_{\vec{\lambda}}(t,x,a) + \int_S V^{\vec{\lambda}}(t,y) q(dy|t,x,a) \tag{5.14}$$

is continuous in $a \in A(t,x)$ by the virtue of [17, Lem.8.3.7], and so $\hat{A}(t,x)$ is closed. Now the compactness of $\hat{A}(t,x)$ follows from this and the compactness of $A(t,x)$; see the discussion immediately after Assumption 4.1.

Let $\hat{K} := \{(t,x,a) : (t,x) \in [0,T]\times S, a \in \hat{A}(t,x)\}$, and $\hat{\mathbb{K}} := \{(t,x,a) : (t,x) \in [0,\infty)\times S, a \in \hat{A}(t,x)\}$. According to Propositions D4 and D5 of [16], it is not hard to see that the set $\hat{\mathbb{K}}$ is a Borel measurable subset of $[0,\infty)\times S\times A$ and $\hat{\mathbb{K}}$ contains the graph of a Borel measurable mapping from $[0,\infty)\times S$ to A.

Now consider a new CTMDP model with n constraints as follows:

$$\hat{\mathbb{M}} := \left\{ S, A, \hat{A}(t,x)(t\geq 0, x\in S), q(dy|t,x,a), (r_k(t,x,a), g_k(x))^n_{k=0} \right\},$$

The corresponding versions of Assumptions 3.1, 3.2, 4.1, and 5.1 are all satisfied by the new model $\hat{\mathbb{M}}$.

Let us consider the space $\hat{\mathcal{U}}$ of performance vectors of the model $\hat{\mathbb{M}}$, and prove

$$\hat{\mathcal{U}} = \mathcal{V} \quad \text{(with } \mathcal{V} \text{ defined by (5.8))}$$

in the following two steps: *i*) $\hat{\mathcal{U}} \subseteq \mathcal{V}$, and *ii*) $\hat{\mathcal{U}} \supseteq \mathcal{V}$.

The proof of *i*): By Theorem 4.1, it suffices to restrict the following arguments to the class of Markov policies. Since each Markov policy in the model $\hat{\mathbb{M}}$ can be regarded as one in the model $\mathbb{M}$, each performance vector in $\hat{\mathcal{U}}$ is also in $\mathcal{U}$ (i.e., $\hat{\mathcal{U}} \subseteq \mathcal{U}$). To further show that each performance vector $\hat{v}$ in $\hat{\mathcal{U}}$ is also on $\mathcal{H}$, let

$$\hat{v} := (V(\hat{\pi}, r_0, g_0), \ldots, V(\hat{\pi}, r_n, g_n)) = (\langle H_0, \eta^{\hat{\pi}} \rangle, \ldots, \langle H_n, \eta^{\hat{\pi}} \rangle)$$

with some Markov policy $\hat{\pi}$ in the model $\hat{\mathbb{M}}$. Then, it follows from Lemma 3.3(a) and (5.4), (5.12)-(5.13) that

$$\begin{aligned}
\sum_{k=0}^{n} \lambda_k \langle H_k, \eta^{\hat{\pi}} \rangle &= \int_{\hat{K}} H^n_{\vec{\lambda}}(t,x,a) \eta^{\hat{\pi}}(dt,dx,da) \\
&= -\int_{\hat{K}} \left[V^{\vec{\lambda}\,\prime}(t,x) + \int_S V^{\vec{\lambda}}(t,y) q(dy|t,x,a) \right] \eta^{\hat{\pi}}(dt,dx,da) \\
&= V^{\vec{\lambda}}(0,\gamma) - \mathbb{E}^{\hat{\pi}}_{\gamma}[V^{\vec{\lambda}}(T,x_T)] \\
&= V^{\vec{\lambda}}(0,\gamma) = \rho^*.
\end{aligned}$$

This means that $\hat{v}$ is on the $\mathcal{H}$, and so $\hat{\mathcal{U}} \subseteq \mathcal{H}$. Hence, we have $\hat{\mathcal{U}} \subseteq \mathcal{U} \cap \mathcal{H} = \mathcal{V}$.

The proof of *ii*): For any fixed $v \in \mathcal{V}$, by Theorem 4.1(b), v can be rewritten as

$$v = (\langle H_0, \eta^{\pi} \rangle, \ldots, \langle H_n, \eta^{\pi} \rangle), \text{ such that } \sum_{k=0}^{n} \lambda_k \langle H_k, \eta^{\pi} \rangle = \rho^*, \text{ for some } \pi \in \Pi^r_m. \tag{5.15}$$

For this Markov policy π, let us define

$$\hat{\Gamma} := \left\{ (t,x) \in [0,T] \times S : \int_A \left(V^{\vec{\lambda}\,\prime}(t,x) + H^n_{\vec{\lambda}}(t,x,a) + \int_S V^{\vec{\lambda}}(t,y) q(dy|t,x,a) \right) \pi(da|t,x) < 0 \right\}$$

Note that $\bar{\eta}^{\pi}(\hat{\Gamma}) = 0$. Indeed, suppose for contradiction that $\bar{\eta}^{\pi}(\hat{\Gamma}) > 0$. Then

$$\begin{aligned}
0 &> \int_S \int_0^T \bar{\eta}^{\pi}(dt,dx) \int_A \left(V^{\vec{\lambda}\,\prime}(t,x) + H^n_{\vec{\lambda}}(t,x,a) + \int_S V^{\vec{\lambda}}(t,y) q(dy|t,x,a) \right) \pi(da|t,x) \\
&= \mathbb{E}^{\pi}_{\gamma}[V^{\vec{\lambda}}(T,x_T)] - V^{\vec{\lambda}}(0,\gamma) + \int_S V(\pi, H^n_{\vec{\lambda}}; 0, x) \gamma(dx) = 0,
\end{aligned}$$

where the last equality is by Lemma 3.3(a) and (5.15). Therefore, $\bar{\eta}^{\pi}(\hat{\Gamma}) = 0$. From this fact and (5.11), we see that the Markov policy $\pi(da|t,x)$ is concentrated on $\hat{A}(t,x)$ for all almost all $(t,x) \in [0,T] \times S$ with respect to the measure $\bar{\eta}^{\pi}(dt,dx)$. Now, there is a set $\zeta \subseteq [0,T] \times S$ of full measure with respect to $\bar{\eta}^{\pi}(dt,dx)$, and a Markov policy $\tilde{\pi}$ satisfying

$$\tilde{\pi}(da|t,x) = \pi(da|t,x)$$

for each $(t,x) \in \zeta$; and

$$\tilde{\pi}(da|t,x) = I_{\{f_{\vec{\lambda}}(t,x)\}}(da)$$

for each $(t,x) \in ([0,T] \times S) \setminus \zeta$. It is clear that this Markov policy $\tilde{\pi}$ is one for the model $\hat{\mathbb{M}}$; see (5.11). For this Markov policy $\tilde{\pi}$, the following relation holds:

$$\eta^{\pi}(dt,dx,da) = \bar{\eta}^{\pi}(dt,dx)\pi(da|t,x) = \bar{\eta}^{\pi}(dt,dx)\tilde{\pi}(da|t,x) = \eta^{\tilde{\pi}}(dt,dx,da),$$

where the last equality is by Theorem 4.1; see its proof. Consequently,

$$(\langle H_0,\eta^{\pi}\rangle,\ldots,\langle H_n,\eta^{\pi}\rangle) = (\langle H_0,\eta^{\tilde{\pi}}\rangle,\ldots,\langle H_n,\eta^{\tilde{\pi}}\rangle) \in \hat{\mathcal{U}}.$$

Consequently, $\mathcal{V} \subseteq \hat{\mathcal{U}}$ because the point $v = (\langle H_0,\eta^{\pi}\rangle,\ldots,\langle H_n,\eta^{\pi}\rangle) \in \mathcal{V}$ is arbitrarily fixed.

Therefore, $\mathcal{V} = \hat{\mathcal{U}}$. Below we legally study the space $\mathcal{V}$ as the space of relevant performance vectors for the model $\hat{\mathbb{M}}$. Since the fixed extreme point v^{ex} of $\mathcal{V}$ is also an extreme point of $\hat{\mathcal{U}} = \mathcal{V}$, and any deterministic Markov policy for the model $\hat{\mathbb{M}}$ is also one for the original model $\mathbb{M}$, to complete the inductive argument, it remains to show that v^{ex} is generated by a deterministic Markov policy for the model $\hat{\mathbb{M}}$.

For the model $\hat{\mathbb{M}}$, a deterministic Markov policy generates the point $v^{ex} = (v_0^{ex}, v_1^{ex}, \ldots, v_n^{ex})$ if and only if it generates $(v_0^{ex}, v_1^{ex}, \ldots, v_{n-1}^{ex})$ because

$$v_n^{ex} = \frac{\rho^* - \sum_{k=0}^{n-1} \lambda_k v_k^{ex}}{\lambda_n}, \tag{5.16}$$

recall that $\lambda_n \neq 0$; see the sentence immediately after (5.6). So, it is equivalent to considering the auxiliary model

$$\hat{\mathbb{M}}' := \{S, A, \hat{A}(t,x)(t \geq 0, x \in S), q(dy|t,x,a), (r_k(t,x,a), g_k(x))_{k=0}^{n-1}\}, \tag{5.17}$$

with only $n-1$ constraints, for which we denote the space of relevant performance vectors by $\hat{\mathcal{U}}'$. For the model $\hat{\mathbb{M}}'$ with $n-1$ constraints, the corresponding versions of Assumptions 3.1, 3.2, 4.1, and 5.1 are all satisfied by this model because so are they by the model $\hat{\mathbb{M}}$ with n constraints. Since $(v_0^{ex}, v_1^{ex}, \ldots, v_n^{ex})$ is an extreme point of $\mathcal{V} = \hat{\mathcal{U}}$, $(v_0^{ex}, v_1^{ex}, \ldots, v_{k-1}^{ex})$ is an extreme point of $\hat{\mathcal{U}}'$, see (5.16). Therefore, by the inductive supposition, the extreme point $(v_0^{ex}, v_1^{ex}, \ldots, v_{n-1}^{ex})$ is generated by a deterministic Markov policy (denoted by f) for the model $\hat{\mathbb{M}}'$. Since f is also in Π_m^r for the model $\mathbb{M}$, it follows from this and (5.16) that the extreme point $v^{ex} = (v_0^{ex}, v_1^{ex}, \ldots, v_n^{ex})$ of $\mathcal{V}$ is generated by the deterministic Markov policy f for the model $\mathbb{M}$. This completes the inductive argument, and (b) is thus proved.

(c) Given parts (a) and (b), the proof of this part of the statement can be similarly proceeded as in the proof of Lemma 9 and Theorem 5 in [15]. □

6 Conclusion

In conclusion, for a constrained CTMDP in a Borel state space, where the performance measures are the expected total rewards over a finite time horizon, under suitable conditions, we showed

the existence of a Markov optimal policy, which is a mixture of $N+1$ deterministic Markov ones, where N is the number of constraints. To this end, we studied the relevant properties of the space of occupation measures and the performance vector spaces.

References

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

[2] Avrachenkov, K., Habachi, O., Piunovskiy, A. and Zhang, Y. (2015). Infinite horizon impulsive optimal control with applications to Internet congestion control. *Int. J. Control* **88**, 703-716.

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

[4] Bertsekas, D., Nedíc, A. and Ozdaglar, A. (2003). *Convex Analysis and Optimization*, Athena Scientific, Belmont.

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

[6] Feinberg, E. and Mandava, M. and Shiryayev, A. (2014). On solutions of Kolmogorovs equations for nonhomogeneous jump Markov processes. *J. Math. Anal. Appl.* **411(1)**, 261-270.

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

[8] Ghosh, M.K. and Saha, S. (2012). Continuous-time controlled jump Markov processes on the finite horizon. *Optimization, Control, and Applications of Stochastic Systems*: 99-109, Birkhäuser, New York.

[9] Guo, X.P. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes.* Springer-Verlag, Berlin.

[10] Guo, X.P. and Hernández-Lerma, O. (2003). Constrained continuous-time Markov controlled processes with discounted criteria. *Stochastic Anal. Appl.* **21**, 379-399.

[11] Guo, X.P., Huang, X.X. and Huang, Y.H. (2015). Finite horizon optimality for continuous-time Markov decision processes with unbounded transition rates. *Adv. Appl. Probab.*, **47**, 1-24.

[12] Guo, X.P., Huang, Y.H. and Song, X.Y. (2012). Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent policies. *SIAM J. Control Optim.* **50**, 23-47.

[13] Guo, X.P. and Song, X.Y. (2011). Discounted continuous-time constrained Markov decision processes in Polish spaces, *Ann. Appl. Probab.* **21**, 2016-2049.

[14] Guo, X.P. and Piunovskiy, A.(2011). Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates, *Math. Oper. Res.* **36**, 105–132.

[15] Guo, X.P., Vykertas, M. and Zhang, Y. (2013). Absorbing continuous-time Markov decision processes with total cost criteria, *Adv. Appl. Probab.* **45**, 490-519.

[16] Hernández-Lerma, O. and Lasserre, J.B. (1996). *Discrete-Time Markov Control Processes: basic optimality criteria*. Springer-Verlag, New York.

[17] Hernández-Lerma, O. and Lasserre, J.B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer-Verlag, New York.

[18] Huang, Y.H., (2015). Finite horizon continuous-time Markov decision processes with mean and variance criteria. Submitted.

[19] Miller, B.L. (1968). Finite state continuous time Markov decision processes with a finite planning horizon. *SIAM. J. Control* **6**,266-280.

[20] Miller, B., Miller, G. and Siemenikhin, K. (2010). Towards the optimal control of Markov chains with constraints. *Automatica* **46**, 1495-1502.

[21] Kitaev, M.Y. and Rykov, V.V.(1995). *Controlled Queueing Systems.* CRC Press, New York.

[22] Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon-Nicodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 235-253.

[23] Piunovskiy, A.B. (1997). *Optimal control of random sequences in problems with constraints*, Kluwer Academic Publishers, Dordrecht.

[24] Piunovskiy, A. (1998). A controlled jump discounted model with constraints. *Theory Probab. Appl.* **42**, 51-71.

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

[26] Pliska, S.R. (1975). Controlled jump processes. *Stochastic Process. Appl.* **3**, 259-282.

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

[28] Yushkevich, A.A. (1977). Controlled Markov models with countable state and continuous time. *Theory Probab. Appl.* **22**, 215–235.

[29] Zhang, L.L. and Guo, X.P. (2008). Constrained continuous-time Markov decision processes with average criteria. *Math. Meth. Oper. Res.* **67**, 323-340.