Optimality of mixed policies for average continuous-time Markov decision processes with constraints

Xianping Guo* and Yi Zhang†

Abstract: This article concerns the average criteria for continuous-time Markov decision processes with N constraints, Under some suitable conditions allowing the transition rates to be possibly unbounded, and the cost rates to be unbounded from both above and below, we establish the following; (a) every extreme point of the space of performance vectors corresponding to the set of stable measures is generated by a deterministic stationary policy; and (b) there exists a mixed optimal policy, where the mixture is over no more than $N + 1$ deterministic stationary policies.

Keywords: Continuous-time Markov decision processes; average criteria; $N + 1$-mixed policy; constrained optimality.

AMS 2000 subject classification: 90C40. 60J25.

1 Introduction

The present paper concerns the average optimality for constrained continuous-time Markov decision processes (CTMDPs).

The average criteria for CTMDPs have been intensively studied; one can find an extensive list of references in the recent monographs [20, 39]. Most of the previous literature focuses on the unconstrained case, and provides conditions for the existence of a deterministic stationary optimal policy out of the more general class of policies. Less literature is available for the constrained problem, where apart from the main long run average cost to be minimized, several other long run averages must be ensured not to exceed their predetermined levels. It is well known that in general the class of deterministic stationary policies is not sufficient for constrained problems; in this case, the standard optimality result is the existence of a randomized stationary policy. In the discrete-time case, every randomized policy can be implemented by performing the randomization procedure at each decision epoch in the standard way. However, as explained by Feinberg in [15, 16], it is impossible to perform the randomization continuously in time. Without a further characterization, it is not clear whether and how a given randomized stationary optimal policy for a CTMDP can be implemented.

Some recent treatments of constrained average CTMDPs include [14, 22, 37, 38] and Chapter 7 of [20]. Only a single constraint is considered in [37, 38] and Chapter 7 of [20]. The model considered in [14] is in finite state and action spaces, for which the author shows the existence of an implementable optimal policy. The model in Chapter 7 of [20] and [37, 38] (resp., [22]) is in a denumerable (resp., possibly uncountable) state space, and the authors show the existence of a randomized stationary optimal policy, whose implementability is left unaddressed. As a fact of matter, in the present literature

*X.P. Guo (E-mail: mcsgxp@mail.sysu.edu.cn). School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, P.R. China

†Y. Zhang (E-mail: yi.zhang@liv.ac.uk). Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK.

we seem not to be aware of any results on this implementability issue for general constrained average CTMDPs in infinite (state and action) spaces. On the other hand, for particular models, one can mention e.g., [35], where an implementable optimal control is provided for a controlled M/M/1 queue with a single constraint.

The main objective of the present paper is to show that there exists an implementable randomized stationary optimal policy for an average CTMDP in Borel spaces with N constraints. Our main contributions are as follows; under some suitable conditions, we show (a) that every extreme point of the space of performance vectors corresponding to the set of stable measures is generated by a deterministic stationary policy (see Theorem 4.1 below); and (b) the optimality of a mixed (randomized stationary) policy, where the mixture is over no more than $N+1$ deterministic stationary policies (see Definition 4.1 and Theorem **??** below). Such an $N+1$-mixed policy can be implemented as follows; one could randomly take a deterministic stationary policy out of the no more than $N+1$ ones according to a specific discrete distribution, and uses the selected deterministic stationary policy to control the process.

To the best of our knowledge, in the previous literature it seems that general results have not been reported about the optimality of mixed policies for constrained CTMDPs in Borel spaces with average criteria, though it has been considered for discrete-time problems and CTMDPs with discounted criteria. One method of establishing the optimality of an $N+1$-mixed policy is based on showing first that each extreme point of the space of occupation or stable measures is generated by a deterministic stationary policy; see, e.g., [2, 3, 6, 16, 21, 34], where [2, 3, 6, 16, 34] deal with discrete-time problems, and [21] considers the discounted criteria for CTMDPs. It seems that establishing this characterization result could be quite involving, especially for general CTMDP models in Borel spaces. Instead, like in [12, 13] and [36] for discrete-time and continuous-time problems with total undiscounted and discounted criteria and [31] focusing on the performance analysis of queueing networks, we pass the average constrained CTMDP problem from the infinite dimensional framework (in the space of measures) to the finite dimensional framework by investigating the space of performance vectors.

The rest of this article is organized as follows. We describe the constrained optimal control problem in Section 2, and then present the preliminaries in Section 3. In Section 4, we formulate and prove the main results. The verifications of all the imposed conditions in this paper are illustrated with an example in Section **??**. The paper is finished with a conclusion in Section **??**. The proofs of the auxiliary results are postponed to the appendix.

2 Optimal control problem

Notation. $I\{\cdot\}$ stands for the indicator function. $\delta_x(\cdot)$ is the Dirac measure concentrated at the point x. $\mathcal{B}(X)$ is the Borel σ-algebra of the metric space X. $\bigvee_{0\le t<s}\mathcal{F}_t$ is the smallest σ-algebra containing all the σ-algebras $\{\mathcal{F}_t, 0\le t<s\}$. $\mathbb{R}_+ := (0,\infty)$. $\mathbb{R}_+^0 := [0,\infty)$. $\mathbb{Z}_+^0 := \{0,1,\dots\}$.

The primitives of a CTMDP are the following elements:

$$\{S, (A(x)\subseteq A, x\in S), q(\cdot|x,a), \gamma\},$$

where

- S (state space): a nonempty Borel space endowed with the Borel σ-algebra $\mathcal{B}(S)$;
- A (action space): a nonempty Borel space endowed with the Borel σ-algebra $\mathcal{B}(A)$;

- $A(x)$ (admissible action spaces given the states $x \in S$): nonempty subsets of A in $\mathcal{B}(A)$ such that the space of admissible state-action pairs

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

is a subset in $\mathcal{B}(S \times A)$ and contains the graph of a (Borel) measurable mapping from S to A;

- $q(dy|x, a)$ (transition rates): a signed kernel on $\mathcal{B}(S)$ given $(x, a) \in \mathbb{K}$, satisfying for each $(x, a) \in \mathbb{K}$, $q(\Gamma_S \setminus \{x\}|x, a) \geq 0$ for all $\Gamma_S \in \mathcal{B}(S)$, $q(S|x, a) = 0$, and for each $x \in S$,

$$\bar{q}_x := \sup_{a \in A(x)} q_x(a) < \infty,$$

where

$$q_x(a) := -q(\{x\}|x, a);$$

- $\gamma(dx)$ (initial distribution): a probability measure on $(S, \mathcal{B}(S))$.

Given the above primitives, one can refer to Kitaev's approach for the construction of the underlying stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P^\pi_\gamma)$ and the controlled process $\{\xi_t, t \geq 0\}$ thereon; see [29, 30]. Below we briefly recall it in order to define the necessary terminologies and notations.

Having joint to $\tilde{\Omega} := (S \times \mathbb{R}_+)^\infty$ all the sequences of the form

$$(x_0,\ \theta_1,\ x_1,\ \ldots,\ \theta_m,\ x_m,\ \infty,\ x_\infty\ ,\infty,\ x_\infty,\ \ldots),$$

where $x_\infty \notin S$ is an isolated point, $x_0 \in S$, $x_l \in S$, $\theta_l \in \mathbb{R}_+$, $1 \leq l \leq m$, and $m \geq 1$, we obtain the sample space $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is the standard Borel σ-algebra. For each $m \geq 0$, define on Ω the measurable mappings T_m, T_∞ and X_m by

$$T_0(\omega) := 0,\ T_m(\omega) := \theta_1 + \theta_2 + \cdots + \theta_m,\ T_\infty(\omega) := \lim_{m\to\infty} T_m(\omega),\ X_m(\omega) := x_m,$$

and the process of interest $\{\xi_t, t \geq 0\}$ by

$$\xi_t(\omega) := \sum_{m\geq 0} I\{T_m \leq t < T_{m+1}\}x_m + I\{T_\infty \leq t\}x_\infty$$

for all $\omega = (x_0,\ \theta_1,\ x_1,\ \ldots,\ \theta_m,\ x_m, \ldots) \in \Omega$, where and below, $0 \cdot x := 0$ for each $x \in S_\infty := S \bigcup \{x_\infty\}$. Let $\mathcal{F}_t := \sigma(\{T_m \leq s, X_m \in \Gamma_S\} : \Gamma_S \in \mathcal{B}(S), s \leq t, m \geq 0)$ for all $t \geq 0$, $A_\infty := A \bigcup \{a_\infty\}$, $A(x_\infty) := \{a_\infty\}$ and $\mathcal{F}_{s-} := \bigvee_{0\leq t<s} \mathcal{F}_t$, where $a_\infty \notin A$ is an isolated point with $q_{x_\infty}(a_\infty) = 0$. The predictable (with respect to $\{\mathcal{F}_t\}_{t\geq 0}$) σ-algebra $\mathbb{P}$ on $\Omega \times \mathbb{R}^0_+$ is given by $\mathbb{P} := \sigma(\Gamma \times \{0\}\ (\Gamma \in \mathcal{F}_0), \Gamma \times (s, \infty)\ (\Gamma \in \mathcal{F}_{s-}))$, see [30, Chap.4] for more details.

Definition 2.1 *A (randomized history-dependent) policy $\pi(\cdot|\omega, t)$ is a $\mathbb{P}$-measurable transition probability function on $(A_\infty, \mathcal{B}(A_\infty))$ concentrated on $A(\xi_{t-}(\omega))$. A policy is called randomized Markov if $\pi(\cdot|\omega, t) = \pi^M(\cdot|\xi_{t-}(\omega), t)$, where $\pi^M(\cdot|x, t)$ is a stochastic kernel on A_∞ given $S_\infty \times \mathbb{R}^0_+$, and $\xi_{t-}(\omega) := \lim_{s\uparrow t} \xi_s(\omega)$. A policy is called randomized stationary if $\pi(\cdot|\omega, t) = \pi^S(\cdot|\xi_{t-}(\omega))$, where $\pi^S(\cdot|x)$ is a stochastic kernel on A_∞ given S_∞. A policy is called deterministic stationary if $\pi(\cdot|\omega, t) = I\{\cdot \ni \phi(\xi_{t-}(\omega))\}$, where $\phi : S_\infty \to A_\infty$ is a measurable mapping such that $\phi(x) \in A(x)$ for all $x \in S_\infty$. Such policies will be denoted as ϕ for simplicity.*

By the way, the term of randomized policies could be also well called relaxed policies as explained in [17, 30]; here we nevertheless follow the practice of calling the relaxed policies "randomized" to be consistent with the majority of the previous literature on this topic [15, 20, 21, 22, 36].

Below we denote by Π_H the class of randomized history-dependent policies, and Π_S the class of randomized stationary policies.

Under each fixed policy $\pi \in \Pi_H$, let us define

$$\nu^\pi(\omega, dt \times \Gamma_S) := \left[\int_A \pi(da|\omega, t) q(\Gamma_S \setminus \{\xi_{t-}(\omega)\}|\xi_{t-}(\omega), a)\right] dt$$

for each $\Gamma_S \in \mathcal{B}(S)$. This random measure is predictable, and such that $\nu^\pi(\omega, \{t\}\times S) = \nu^\pi(\omega, [T_\infty, \infty)\times S) = 0$, see [29, 30]. Therefore, there exists a unique probability measure P^π_γ such that $P^\pi_\gamma(\xi_0 \in dx) = \gamma(dx)$, and with respect to P^π_γ, ν^π is the dual predictable projection of the random measure of the marked point process $(T_m,\ X_m)$ with its internal history, see [28, 29, 30]. In what follows, when $\gamma(\cdot)$ is a Dirac measure $\delta_x(\cdot)$ concentrated at $x \in S$, we use the degenerated notation P^π_x. Expectations with respect to P^π_γ and P^π_x are denoted as E^π_γ and E^π_x, respectively.

The following condition guarantees the nonexplosiveness of the controlled process under each policy; see more comments on this after the condition.

Condition 2.1 *There exist a continuous $[1, \infty)$-valued function w on S and constants $\rho \in \mathbb{R}$, $b \geq 0$ such that*

(a) $\bigcup_{l=0}^\infty S_l = S$, $\lim_{l\to\infty} \inf_{x\in S\setminus S_l} w(x) = \infty$ *for an increasing sequence of measurable subsets $S_l \subseteq S$.*

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

(c) *For any $l \in \mathbb{Z}^0_+$, $\sup_{x\in S_l} \bar{q}_x < \infty$, where the sets S_l are from part (a) of this condition.*

Here and below we formally adopt the convention that the infimum taken over the empty set is ∞.

Condition 2.1 guarantees that the controlled process $\{\xi_t, t \geq 0\}$ is nonexplosive under each policy π, i.e.,

$$P^\pi_x(T_\infty = \infty) = 1,\ \forall\ x \in S;$$

see Lemma 2.1. The origin of Condition 2.1 is [7] by M. Chen, where it is shown to be sufficient for the nonexplosiveness for the (uncontrolled) time-homogeneous Markov pure jump process. Recently, when the state space is denumerable, F. Spieksma [44] showed that this condition is actually also necessary for the nonexplosiveness; see also the discussions in the recent paper by M. Chen [9]. It was brought to our attention by a referee that sufficient conditions for the nonexplosiveness of the time-inhomogeneous Markov pure jump process were also provided in the less known Chinese literature; see J. Zheng [47].

For the optimal control problem (1) considered below, we will show that one can concentrate on stationary policies that induce invariant probabilities; see Proposition 3.2. That result could fail to hold if the process is explosive (so that in particular Condition 2.1 is violated); see Example 3.1 below.

The next lemma comes from [36].

Lemma 2.1 *Suppose Condition 2.1 is satisfied, where $\rho \neq 0$. Then, the following assertions hold for each policy π, $x \in S$ and $t \geq 0$.*

(a) $P^\pi_x(T_\infty = \infty) = 1$.

(b) $E^\pi_x[w(\xi_t)] \leq e^{-\rho t}w(x) + \frac{b}{\rho}(1 - e^{-\rho t})$.

Let $c_i(x,a)$, $i=0,1,\ldots,N$, be measurable (real-valued) functions on $\mathbb{K}$, representing the cost rates, and $d_j \in \mathbb{R}, j=1,2,\ldots,N$, be the predetermined constraint constants. Introduce

$$V(\gamma,\pi,g) \;:=\; \overline{\lim_{T\to\infty}}\frac{1}{T}E_\gamma^\pi\left[\int_0^T\int_A g(\xi_t,a)\pi(da|\omega,t)dt\right]$$

for each measurable function g on $\mathbb{K}$ (whenever the right hand side of the above is well defined), whereas if the initial distribution γ is a Dirac measure concentrated at a state $x\in S$, $V(\gamma,\pi,g)$ is written as $V(x,\pi,g)$. Then, the constrained average CTMDP optimal control problem under consideration reads

$$\begin{aligned} V(\gamma,\pi,c_0) &\to \min_{\pi\in\Pi_H} \\ \text{subject to } V(\gamma,\pi,c_j) &\le d_j,\ j=1,2,\ldots,N. \end{aligned} \tag{1}$$

The next statement immediately follows from Lemma 2.1.

Lemma 2.2 *Suppose Condition 2.1 is satisfied. If there exists a constant $M\ge 0$ such that*

$$\sup_{a\in A(x)}|c_i(x,a)|\le Mw(x)\ \forall\ i=0,1,\ldots,N,$$

$\int_S w(y)\gamma(dy)<\infty$, and $\rho>0$, where w and ρ come from Condition 2.1, then under each policy $\pi\in\Pi_H$,

$$\overline{\lim_{T\to\infty}}\frac{1}{T}E_\gamma^\pi\left[\int_0^T\int_A |c_i(\xi_t,a)|\pi(da|\omega,t)dt\right]\le\frac{bM}{\rho}\ \forall\ i=0,1,\ldots,N.$$

Definition 2.2 *A policy satisfying all the N constraints in problem (1) is called feasible. A feasible policy solving problem (1) is called (constrained average) optimal.*

Throughout this article, to avoid trivial cases, we take the following assumption as granted, which is not mentioned explicitly below.

Assumption 2.1 *There exists at least one feasible policy to problem (1).*

3 Preliminaries

Given any probability measure η on $\mathbb{K}$, one can disintegrate it with respect to its marginal $\eta(dx,A)$ to get a unique (in the almost sure sense) stochastic kernel $\pi_\eta(da|x)$, defining a (possibly randomized) stationary policy denoted as π_η, so that

$$\eta(dx,da)=\eta(dx,A)\pi_\eta(da|x);$$

see [23]. Here and below when it simplifies the notations, we may freely regard such η as measures on $S\times A$ concentrated on $\mathbb{K}$.

Definition 3.1 *Suppose Condition 2.1 is satisfied, where $\rho>0$, and $\sup_{x\in S}\frac{\overline{q}_x}{w(x)}<\infty$. A probability measure η on $\mathbb{K}$ is said to be stable if*

$$\int_S w(x)\eta(dx,A)\le 1+\frac{b}{\rho} \tag{2}$$

and

$$\int_S\int_A q(\Gamma_S|x,a)\pi_\eta(da|x)\eta(dx,A)=0 \tag{3}$$

for all $\Gamma_S\in\mathcal{B}(S)$. On this occasion, the underlying stationary policy π_η is said to be stable, too.

We denote by $\mathcal{D}$ the collection of such stable probability measures on $\mathbb{K}$, and by Π_{Stable} the class of stable policies. Then it holds that $\Pi_{Stable} \subseteq \Pi_S$. Relation (3) implies that $\eta(dx, A)$ is an invariant probability for $\int_A q(\cdot|x,a)\pi_\eta(da|x)$; see [8].

Definition 3.2 *Let $f \geq 1$ be a measurable function on S.*

(a) *A probability measure μ on $\mathbb{K}$ (resp., S) is said to be f-bounded if*

$$\int_{\mathbb{K}} f(x)\mu(dx, da) < \infty \ (resp., \ \int_S f(x)\mu(dx) < \infty).$$

The collection of f-bounded probability measures on $\mathbb{K}$ (resp., S) is denoted by $\mathcal{P}_f(\mathbb{K})$ (resp., $\mathcal{P}_f(S)$).

(b) *A measurable function u on $\mathbb{K}$ (resp., S) is said to be f-bounded if*

$$\sup_{x\in S} \frac{\sup_{a\in A(x)} |u(x,a)|}{f(x)} < \infty \ (resp., \ \sup_{x\in S} \frac{|u(x)|}{f(x)} < \infty).$$

(c) *The f-weak topology on $\mathcal{P}_f(\mathbb{K})$ is the weakest topology such that for each f-bounded continuous function u on $\mathbb{K}$, $\int_{\mathbb{K}} u(x,a)\mu(dx,da)$ is continuous in $\mu \in \mathcal{P}_f(\mathbb{K})$. This topology is denoted by $\tau(\mathcal{P}_f(\mathbb{K}))$.*

The f-weak topology on other Borel spaces is similarly defined. The convergence in the f-weak topology is denoted by "$\xrightarrow{f}$".

There is a one-to-one correspondence T_f between $\mathcal{P}_1(\mathbb{K})$ and $\mathcal{P}_f(\mathbb{K})$, where $f \geq 1$ is a fixed continuous function on S. Indeed, for each $\mu \in \mathcal{P}_f(\mathbb{K})$, one can define $\tilde{\mu} \in \mathcal{P}_1(\mathbb{K})$ by

$$\tilde{\mu}(\Gamma) = T_f(\mu)(\Gamma) := \frac{\int_\Gamma f(x)\mu(dx,da)}{\int_{\mathbb{K}} f(x)\mu(dx,da)} \ \forall \ \Gamma \in \mathcal{B}(\mathbb{K}); \tag{4}$$

and given any $\tilde{\mu} \in \mathcal{P}_1(\mathbb{K})$, one can define $\mu \in \mathcal{P}_f(\mathbb{K})$ by

$$\mu(\Gamma) := T_f^{-1}(\tilde{\mu})(\Gamma) = \frac{\int_\Gamma \frac{1}{f(x)}\tilde{\mu}(dx,da)}{\int_{\mathbb{K}} \frac{1}{f(x)}\tilde{\mu}(dx,da)} \ \forall \ \Gamma \in \mathcal{B}(\mathbb{K}). \tag{5}$$

The next lemma comes from [36, Lem.3.4, Rem.3].

Lemma 3.1 *Suppose a continuous function $f \geq 1$ on S is fixed. Then the two topological spaces $(\mathcal{P}_f(\mathbb{K}), \tau(\mathcal{P}_f(\mathbb{K})))$ and $(\mathcal{P}_1(\mathbb{K}), \tau(\mathcal{P}_1(\mathbb{K})))$ are homeomorphic, with the mapping T_f defined by (4) being a homeomorphism. In particular, $(\mathcal{P}_f(\mathbb{K}), \tau(\mathcal{P}_f(\mathbb{K})))$ is metrizable because so is $(\mathcal{P}_1(\mathbb{K}), \tau(\mathcal{P}_1(\mathbb{K})))$.*

To show the compactness of the set $\mathcal{D}$ in $(\mathcal{P}_{w'}(\mathbb{K}), \tau(\mathcal{P}_{w'}(\mathbb{K})))$ and the existence of an optimal policy, we impose the following condition; see more discussions on the various consequences of the imposed condition in the remarks following it.

Condition 3.1 *Let w be as in Condition 2.1.*

(a) *$\int_S g(y)q(dy|x,a)$ is continuous on $\mathbb{K}$ for each bounded continuous function $g(\cdot)$ on S.*

(b) *There exists a continuous moment function $w' \geq 1$ on S and a constant $M' \geq 0$ such that $\overline{q}_x \leq M'w'(x)$ and $\sup_{a\in A(x)} |c_i(x,a)| \leq M'w'(x)$ for all $x \in S$ and $i = 0, 1, \ldots, N$.*

(c) *There exists an increasing sequence of compact sets $K_m \uparrow \mathbb{K}$ such that*

$$\lim_{m\to\infty} \inf_{(x,a)\in\mathbb{K}\setminus K_m} \frac{w(x)}{w'(x)} = \infty.$$

(d) *$\rho > 0$ and $\int_S w(y)\gamma(dy) < \infty$, where the constant ρ is as in Condition 2.1.*

In case $\mathbb{K}$ is itself compact, for verifying this condition one could take $w' \geq 1$ as any w-bounded continuous function because of the convention that any infimum taken over the empty set is put ∞.

Remark 3.1 *It follows from [36, Lem.3.10] that Condition 3.1(c) implies that $A(x)$ is compact for any $x \in S$.*

Remark 3.2 *(a) Under Conditions 2.1 and 3.1(b,c), there exists a compact set $K_m \subseteq \mathbb{K}$ with a large enough index m such that*

$$\sup_{(x,a)\in\mathbb{K}\setminus K_m} \frac{w'(x)}{w(x)} = \frac{1}{\inf_{(x,a)\in\mathbb{K}\setminus K_m} \frac{w(x)}{w'(x)}} < \infty; \quad \sup_{(x,a)\in K_m} \frac{w'(x)}{w(x)} < \infty,$$

so that the function w' is w-bounded. This fact also guarantees that the space of stable measures $\mathcal{D}$ is a subset of $\mathcal{P}_{w'}(\mathbb{K})$.
(b) By [23, Rem.5.7.5, p.115], Condition 3.1(c) is satisfied if the following holds: (i) the set $\{x \in S : A(x) \subseteq G\}$ is open in S for every open set $G \subseteq A$; (ii) both S and A are σ-compact; and for each $\epsilon > 0$, there exists a compact set $S_\epsilon \subseteq S$ such that $\frac{w(x)}{w'(x)} \geq \epsilon$ for all $x \in S \setminus S_\epsilon$; and (iii) $A(x)$ is compact for each $x \in S$.

Remark 3.3 *Let $t_0 > 0$ be fixed. Condition 3.1 together with Condition 2.1 guarantees the uniform integrability with respect to the cost rates c_i and the precompactness properties of the family $\{\eta_t^\pi,\ t \geq t_0\}$ of empirical measures in $\mathcal{P}_{w'}(\mathbb{K})$ endowed with the w'-weak topology, where for each $t > 0$ and policy π,*

$$\eta_t^\pi(\gamma, dx, da) := \frac{1}{t} E_\gamma^\pi \left[\int_0^t I\{\xi_s \in dx\} \pi(da|\omega, s) ds \right]. \tag{6}$$

(If $\gamma(dy) = I\{z \in dy\}$ for some $z \in S$, then we write $\eta_t^\pi(z, dx, da)$.) Similar properties for empirical measures also play an important role in the investigations of discrete-time problems; see e.g., Altman and Shwartz [2] and Altman [3]. In greater detail, it follows from Lemma 2.1(b) that

$$\sup_{t\geq t_0} \int_{\mathbb{K}} \frac{w(x)}{w'(x)} T_{w'}(\eta_t^\pi)(dx, da) < \infty$$

(Condition 3.1(d) in particular guarantees the inequality). This fact, according to Theorem 12.2.15 of [24], implies that the family $\{T_{w'}(\eta_t^\pi),\ t \geq t_0\}$ is tight, and thus precompact in $\mathcal{P}_1(\mathbb{K})$ by the Prokhorov theorem; see Theorem 12.2.16 of [24]. It remains to apply Lemma 3.1. (The same reasoning is also used in the proof of Proposition 3.1 below to show that the space of stable measures $\mathcal{D}$ is precompact in $\mathcal{P}_{w'}(\mathbb{K})$ endowed with the w'-weak topology. Then Condition 3.1(a) guarantees that $\mathcal{D}$ is closed and thus compact in $\mathcal{P}_{w'}(\mathbb{K})$ endowed with the w'-weak topology.)

It also follows from the tightness of $\{T_{w'}(\eta_t^\pi),\ t \geq t_0\}$ and the fact of $\sup_{t\geq t_0} \int_{\mathbb{K}} w'(y)\eta_t^\pi(\gamma, dy, da) < \infty$ that under each policy π, $\{\eta_t^\pi,\ t \geq t_0\}$ is uniformly integrable with respect to c_i, $i = 0, 1, \ldots, N$ (see Definition A.4 in Altman [3]); recall the fact that the cost rates c_i, $i = 0, 1, \ldots, N$ are w'-bounded under Condition 3.1.

Proposition 3.1 *Suppose Conditions 2.1 and 3.1 are satisfied. Then the space of stable measures $\mathcal{D}$ is nonempty, convex and compact in $(\mathcal{P}_{w'}(\mathbb{K}), \tau(\mathcal{P}_{w'}(\mathbb{K})))$.*

Proof. See the appendix. □

Condition 3.2 *For each stable policy π_η corresponding to a stable measure $\eta \in \mathcal{D}$, it holds that*

$$V(\gamma, \pi_\eta, c_i) = \lim_{T\to\infty} \frac{1}{T} \int_0^T E_\gamma^{\pi_\eta} \left[\int_A c_i(\xi_t, a) \pi_\eta(da|\xi_t) \right] dt \quad = \quad \int_{\mathbb{K}} c_i(y, a) \eta(dy, da)$$

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

Remark 3.4 *Suppose Conditions 2.1 and 3.1 hold. Then Condition 3.2 is satisfied if under each stable policy π, the controlled process is positive Harris recurrent. (Remember, a stable policy π is stationary.) Indeed, in this case, under each stable policy π, there is a unique invariant probability η^π, and by Theorem 1 of [18], for each $z \in S$, as $t \to \infty$, $\eta_t^\pi(z, dx, A)$ converges to the unique invariant probability $\eta^\pi(dx)$ setwise, where η_t^π is defined by (6). It follows from Remark 3.3 that necessarily $\eta_t^\pi(z, dx, A)$ converges to $\eta^\pi(dx)$ in the w'-weak topology for each $z \in S$, and furthermore, $\int_S \eta_t^\pi(z, dx, A) w'(x) < \infty$ and $\int_S w'(y) \eta^\pi(dy) < \infty$. Now according to Theorem 2.4 of [42], Condition 3.2 is satisfied; recall that the functions $x \in S \to \int_A c_i(x, a)\pi(da|x)$, $i = 0, 1, \ldots, N$ are w'-bounded.*

In particular, Condition 3.2 is satisfied by the finite unichain model, which means that the state and action spaces are both finite, and under each deterministic stationary (and thus each randomized stationary) policy, the controlled process admits a unique positive recurrent class plus a possibly empty set of transient states.

Condition 3.2 implicitly reduces to at least the uniqueness of the invariant probability for the controlled process under each stable policy, although if the cost rates are constant, then Condition 3.2 becomes trivial, without requiring any properties to be exhibited by the controlled process.

Lemma 3.2 *Suppose Conditions 2.1, 3.1 and 3.2 are satisfied, and the functions $c_i, i = 0, 1, \ldots, N$, are all lower semicontinuous on $\mathbb{K}$. Then,*

(a) *for any $i = 0, 1, \ldots, N$, $\int_{\mathbb{K}} c_i(x, a)\eta(dx, da)$ is lower semicontinuous in $\eta \in \mathcal{P}_{w'}(\mathbb{K})$ (equipped with the w'-weak topology);*

(b) *for each policy π, there exists a stable measure $\eta \in \mathcal{D}$ with an associated stable policy π_η such that*

$$V(\gamma, \pi_\eta, c_i) \le V(\gamma, \pi, c_i), \ i = 0, 1, \ldots, N.$$

Proof. See the appendix. □

As a consequence of Lemma 3.2, for problem (1) it suffices to consider the class of stable policies, and problem (1) can be reformulated as

$$\begin{aligned} & \int_{\mathbb{K}} c_0(x, a)\eta(dx, da) \to \min_{\eta\in\mathcal{D}} \\ s.t. \quad & \int_{\mathbb{K}} c_j(x, a)\eta(dx, da) \le d_j, j = 1, 2, \ldots, N. \end{aligned} \tag{7}$$

Proposition 3.2 *Suppose Conditions 2.1, 3.1 and 3.2 are satisfied, and $c_i(x, a)$ $(i = 0, \ldots, N)$ are all lower semicontinuous on $\mathbb{K}$. Then, there is an optimal solution to problem (7), and thus a stable optimal policy exists for the constrained average CTMDP problem (1).*

Proof. See the appendix below. □

The following example shows that if Condition 2.1 is not satisfied, then there might not be any optimal stationary policy that induces an invariant probability (c.f. Proposition 3.2).

Example 3.1 *Let $S = \{0, \pm 1, \pm 2, \dots\}$, $A = \{a_1, a_2\} = A(0)$, $A(x) = A$ for all $0 \neq x \in S$. Let $0 < \mu < \lambda < 2\mu$ be fixed constants such that $\lambda + \mu = 1$. Consider the transition rate given by*

$$q_0(a_1) = q(\{1\}|0, a_1) = \lambda = q_0(a_2) = q(\{-1\}|0, a_2);$$
$$q_x(a) = q(\{x-1\}|x, a) = 1, \ \forall x \in \{-1, -2, \dots\}, a \in A;$$
$$q(\{x+1\}|x, a) = \lambda 2^x, \ q(\{x-1\}|x, a) = \mu 2^x, \ q_x(a) = 2^x, \ \forall x \in \{1, 2, \dots\}, a \in A.$$

Let us fix a single cost rate given by

$$c_0(x, a) = 0, \forall \ x \in \{0, 1, \dots\}, a \in A;$$
$$c_0(x, a) = -1, \forall \ x \in \{-1, -2, \dots\}, a \in A,$$

We introduce the notation

$$\pi(\{a_1\}|0) = \gamma \in [0, 1], \ \pi(\{a_2\}|0) = 1 - \gamma.$$

Note that the process is controlled only at the state 0, and so a stationary policy $\pi(da|x)$ is fully specified by the constant $\gamma \in [0, 1]$.

Under the stationary policy π with $\gamma \in [0, 1)$ being arbitrarily fixed, it is evident that there does not exist any invariant probability for $\int_A q(\cdot|x, a)\pi(da|x)$. In other words, any stationary policy $\pi(da|x)$ specified by some $\gamma \in [0, 1)$ does not induce an invariant probability.

When $\gamma = 1$, the stationary policy becomes deterministic, under which there is a unique invariant probability p given by

$$p(\{x\}) = 0, \ \forall x = -1, -2, \dots;$$
$$p(\{x\}) = \left(1 - \frac{\lambda}{2\mu}\right)\left(\frac{\lambda}{2\mu}\right)^x, \ \forall \ x = 0, 1, 2, \dots.$$

Therefore, the deterministic stationary policy given by $f_0(0) = a_1$ is the unique stationary policy that induces an invariant probability.

It is obvious that the deterministic stationary policy $f^*(0) = a_2$ is optimal with

$$V(0, f^*, c_0) = -1 < V(0, f_0, c_0) = 0,$$

which thus strictly outperforms the unique stationary policy f_0 that induces an invariant probability.

In the previous example (c.f. [4]), under the policy f_0, the controlled process is explosive. To avoid the explosiveness, we imposed Condition 2.1. □

We finish this section with some additional notations, conditions and technical results, which are to be used in the next section. For each $x \in S$, let $\hat{A}(x) \subseteq A(x)$ be an arbitrarily fixed nonempty compact subset of A such that $\hat{K} := \{(x, a) : x \in S, a \in \hat{A}(x)\}$ is measurable and contains the graph of a measurable mapping from S to A. We consider the so-called $\hat{A}$-CTMDP model $\{S, (\hat{A}(x) \subseteq A, x \in S), q(\cdot|x, a), \gamma\}$, which is a specific sub-model of $\{S, (A(x) \subseteq A, x \in S), q(\cdot|x, a), \gamma\}$. Let $\alpha > 0$ be arbitrarily fixed. Then we define the following discounted criterion for the $\hat{A}$-CTMDP model (restricted to the class of deterministic stationary policies):

$$V^{\alpha}_{c^{\vec{\lambda}}}(\hat{\phi}, x) := E^{\hat{\phi}}_x \left[\int_0^{\infty} e^{-\alpha t} c^{\vec{\lambda}}(\xi_t, \hat{\phi}(\xi_t))dt\right], \tag{8}$$

with the value function being denoted by

$$V^{\alpha}_{c^{\vec{\lambda}}}(x) := \inf_{\hat{\phi}} V^{\alpha}_{c^{\vec{\lambda}}}(\hat{\phi}, x), \tag{9}$$

where the infimum is taken over the class of deterministic stationary policies $\hat{\phi}$ for the $\hat{A}$-CTMDP model, and

$$c^{\vec{\lambda}}(x,a) := \sum_{i=0}^{N} \lambda_i c_i(x,a),$$

with $\vec{\lambda} := (\lambda_0, \cdots, \lambda_N) \in \mathbb{R}^{N+1}$. Clearly, $V^{\alpha}_{c^{\vec{\lambda}}}(\hat{\phi}, x)$ and $V^{\alpha}_{c^{\vec{\lambda}}}(x)$ depend on $\hat{A}(\cdot)$, but we do not indicate this dependence in the denotations for brevity.

Condition 3.3 *Let w and w' be as in Condition 3.1.*

(a) *The functions $c_i, i = 0, 1, \ldots, N$, are continuous on $\mathbb{K}$.*

(b) *There exist constants $\overline{M} \in \mathbb{R}$ and $\rho' > 0, b' \geq 0$ such that for each $x \in S, a \in A(x)$,*

$$(\overline{q}_x + 1)w'(x) \;\leq\; \overline{M}w(x); \;\; \int_S q(dy|x,a)w'(y) \leq -\rho' w'(x) + b'.$$

In case the state space S is denumerable and the model is unichain, the previous conditions have the following consequences.

Proposition 3.3 *Suppose Conditions 2.1, 3.1, and 3.3 are satisfied, the state space S is denumerable, and each deterministic stationary policy is unichain. Then the following assertions holds.*

(a) *For each $\vec{\lambda} \in \mathbb{R}^{N+1}$ and $\hat{A}$-CTMDP model, there exist constants $L_{c_{\vec{\lambda}}}, \alpha_{c_{\vec{\lambda}}} > 0$ and some state $x_{c_{\vec{\lambda}}} \in S$ such that*

$$|V^{\alpha}_{c^{\vec{\lambda}}}(x) - V^{\alpha}_{c^{\vec{\lambda}}}(x_{c_{\vec{\lambda}}})| \leq L_{c_{\vec{\lambda}}} w'(x) \tag{10}$$

for all $x \in S$ and $\alpha \in (0, \alpha_{c_{\vec{\lambda}}})$. (Here the $L_{c_{\vec{\lambda}}}, \alpha_{c_{\vec{\lambda}}} > 0$ and $x_{c_{\vec{\lambda}}} \in S$ are possibly dependent on $\hat{A}(\cdot)$ and $c_{\vec{\lambda}}$.)

(b) *Every deterministic stationary policy is stable. (This holds without requiring apriori Condition 3.3 to hold, or that each deterministic stationary policy is unichain.)*

(c) *Condition 3.2 is satisfied.*

Proof. As for part (b), as in Remark 3.3, one can see that for each $t_0 > 0$, the family $\{\eta^f_t,\ t \geq t_0\}$ is tight for each initial state $z \in S$. As a result, the controlled process (under each deterministic stationary policy) is bounded in probability on average, and now part (b) follows from Theorem 3.1 of [32]. The reasoning in the proof of Theorem 3.13 in [41] applies to show that under the conditions of the statement, the A-CTMDP model (and thus each of the $\hat{A}$-CTMDP model) is uniformly w'-exponentially ergodic with respect to all randomized stationary policies. Following from this, parts (a) and (c) immediately hold; for part (a), further see the reasoning in the proof of Lemma 7.7 of [20]. □

The proof of Proposition 3.3 (see part (a) therein) makes use of the fact that under the conditions therein, the controlled process in a denumerable state space is uniformly w'-exponentially ergodic with respect to all stationary policies; see Theorem 3.13 of Prieto-Rumeau and Hernández-Lerma [41],

whose proof is based on the relevant results for denumerable state discrete-time models in Dekker et al [10]; see also Spieksma [43]. Since this extension to the case of an uncountable state space is not yet immediate, we impose the assertions of Proposition 3.3 to hold as in the following condition (for the case of an uncountable state space).

Condition 3.4 **(a)** *For each bounded measurable function g on S, $\int_S g(y)q(dy|x,a)$ is continuous in $a \in A(x)$ for each fixed $x \in S$.*

(b) *Parts (a), (b) and (c) of Proposition 3.3 hold. Furthermore, for each $\eta \in \mathcal{D}$, if $\eta(Z,A) = 0$ for some $Z \in \mathcal{B}(S)$, then $\eta'(Z,A) = 0$ for all $\eta' \in \mathcal{D}$.*

As mentioned in the above, for the verification of the above condition, the validity of (a) of Proposition 3.3 (see (10)) is the least transparent; it is satisfied if the controlled model is uniformly w'-exponentially ergodic, for which some sufficient conditions (of the stochastic monotonicity type) in the uncountable state space case are given in [19], see also [46]. For the last part of Condition 3.4(b), we mention that it is not needed if the state space is denumerable, or in fact, if the functions u_1^* and u_2^* in Lemma 3.3 below coincide.

The next result can be useful in verifying the last part of Condition 3.4(b). Its proof has been omitted.

Proposition 3.4 *Suppose Conditions 2.1 and 3.1 are satisfied. If there exists a non-trivial σ-finite measure ν on S and a positive-valued function $g(x,a,y) > 0$ on $\mathbb{K} \times S$ such that*

$$q(D|x,a) = \int_D g(x,a,y)\nu(dy) \ \forall \ D \in \mathcal{B}(S), \ x \notin D, \ a \in A(x).$$

then Condition 3.4(b) is satisfied.

Remark 3.5 *Under Conditions 2.1, 3.1(b,c), 3.3 and 3.4(a), for each $x \in S$, $\int_S u(y)q(dy|x,a)$ is continuous in $a \in A(x)$ for each measurable function u satisfying $\sup_{x\in S} \frac{|u(y)|}{w'(y)} < \infty$. This follows from the reasoning in the proof of Corollary 2.6 of [40].*

Finally we present the following statement about unconstrained average CTMDPs (c.f. [19]), which serves the proof in the next section.

Lemma 3.3 *Suppose Conditions 2.1, 3.1(b,c), 3.3, and 3.4 are satisfied. The following assertions hold.*

(a) *For each $\vec{\lambda} \in \mathbb{R}^{N+1}$, there exist a constant $v^*(\vec{\lambda}) \in \mathbb{R}$, w'-bounded measurable functions u_1^*, u_2^* on S and deterministic stationary policies φ^*, ψ^*, all of which are possibly $\vec{\lambda}$-dependent, such that for all $x \in S$,*

$$c^{\vec{\lambda}}(x,\varphi^*(x)) + \int_S u_1^*(y)q(dy|x,\varphi^*(x)) = \inf_{a\in A(x)} \left\{ c^{\vec{\lambda}}(x,a) + \int_S u_1^*(y)q(dy|x,a) \right\} \le v^*(\vec{\lambda});$$

$$v^*(\vec{\lambda}) \le \inf_{a\in A(x)} \left\{ c^{\vec{\lambda}}(x,a) + \int_S u_2^*(y)q(dy|x,a) \right\} = c^{\vec{\lambda}}(x,\psi^*(x)) + \int_S u_2^*(y)q(dy|x,\psi^*(x)).$$

(b) $\inf_{\pi\in\Pi_{Stable}} V(\gamma,\pi,c^{\vec{\lambda}}) = V(\gamma,\varphi^*,c^{\vec{\lambda}}) = v^*(\vec{\lambda})$, *where φ^* is as in part (a).*

(c) *If a stable policy π satisfies $V(\gamma,\pi,c^{\vec{\lambda}}) = v^*(\vec{\lambda})$, then there exists a measurable subset $S_\pi^{\vec{\lambda}} \subseteq S$ (depending on π and $\vec{\lambda}$) such that*

$$\eta^\pi(S_\pi^{\vec{\lambda}}, A) = 1,$$

and

$$\pi(B(x)|x) = 1 \ \forall \ x \in S_\pi^{\vec{\lambda}},$$

where

$$B(x) := \left\{ a \in A(x) : c^{\vec{\lambda}}(x,a) + \int_S u_2^*(y) q(dy|x,a) = v^*(\vec{\lambda}) \right\},$$

and $\eta^\pi(dx,da)$ denotes the stable measure corresponding to π.

(d) *In case the state space S is denumerable, u_1^* and u_2^* from part (a) coincide, and one can take $\psi^* = \varphi^*$; $B(x)$ from part (c) is nonempty for each $x \in S$.*

Proof. See the appendix below. □

4 Main results

Definition 4.1 *A stable policy (with respect to a stable measure η) is called mixed over a class of $m+1$ deterministic stationary stable policies $\varphi_l, l = 0,1,2,\ldots,m$, if*

$$\eta(dx,da) = \sum_{l=0}^{m} b_l \eta_l(dx,da),$$

where η_l are the stable measures corresponding to φ_l, and the nonnegative constants b_l satisfy $\sum_{l=0}^m b_l = 1$.

Denote by

$$\mathcal{V} := \{(V(\gamma,\pi,c_0), V(\gamma,\pi,c_1), \ldots, V(\gamma,\pi,c_N)) : \ \pi \in \Pi_{Stable}\} \subseteq \mathbb{R}^{N+1} \tag{11}$$

the space of (relevant) performance vectors (generated by stable policies) for the original average CTMDP model $\{S, A, A(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}$.

Denote by

$$\mathcal{V} := \{(V(\gamma,\pi,c_0), V(\gamma,\pi,c_1), \ldots, V(\gamma,\pi,c_N)) : \ \pi \in \Pi_{Stable}\} \subseteq \mathbb{R}^{N+1} \tag{12}$$

the space of (relevant) performance vectors (generated by stable policies) for the original average CTMDP model $\{S, A, A(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}$.

Theorem 4.1 *Suppose Conditions 2.1, 3.1, and 3.3 are satisfied. Consider the following two situations:*

(a) the state space S is denumerable and the model is unichain;

(b) the state space S is uncountable (Borel), and additionally Condition 3.4 is satisfied.

In either case, the space of performance vectors $\mathcal{V}$ is nonempty, compact and convex, and each extreme point of $\mathcal{V}$ (there exists at least one), say v^{ex}, is generated by a deterministic stationary policy, say φ, i.e.,

$$v^{ex} = (V(\gamma, \varphi, c_0), V(\gamma, \varphi, c_1), \ldots, V(\gamma, \varphi, c_N)).$$

Proof. It is clear that

$$\mathcal{V} = \Phi(\mathcal{D}) := \left\{ \left(\int_{\mathbb{K}} c_0(x,a)\eta(dx,da), \int_{\mathbb{K}} c_1(x,a)\eta(dx,da), \ldots, \int_{\mathbb{K}} c_N(x,a)\eta(dx,da) \right), \eta \in \mathcal{D} \right\},$$

where, under the conditions of the theorem, Φ is a w'-continuous mapping from $\mathcal{D}$ to $\mathcal{V}$ equipped with the usual Euclidean topology. Therefore, by [1, Thm.2.34], $\mathcal{V}$ is nonempty, convex and compact, because so is $\mathcal{D}$, according to Proposition 3.1. So by [1, Cor.7.66], $\mathcal{V}$ admits at least one extreme point, say v^{ex}. Below we prove that any given extreme point v^{ex} of $\mathcal{V}$ is generated by a deterministic stationary policy by induction with respect to the number of constraints N.

Consider the case of $N = 0$, i.e., consider an unconstrained CTMDP model satisfying Conditions 2.1, 3.1, 3.2, and 3.3 in case the state space S is denumerable, and additionally Condition 3.4 in case S is uncountable. Then by the convexity and compactness of $\mathcal{V}$ proved above, $\mathcal{V} \subseteq \mathbb{R}$ is a bounded closed interval, and the two extreme points of $\mathcal{V}$, denoted 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 average CTMDP problems

$$V(\gamma, \pi, c_0) \quad \to \quad \min_{\pi \in \Pi_{Stable}} \tag{13}$$

and

$$V(\gamma, \pi, c_0) \quad \to \quad \max_{\pi \in \Pi_{Stable}}, \tag{14}$$

respectively. For problem (13), by Lemma 3.3, there is a deterministic stationary policy, say φ_1, such that

$$v_{min} = \inf_{\pi \in \Pi_{Stable}} V(\gamma, \pi, c_0) = V(\gamma, \varphi_1, c_0).$$

For problem (14), especially due to the continuity of $c_0(x,a)$, its optimal policy is given by the optimal solution to the problem $V(\gamma, \pi, -c_0) \to \min_{\pi \in \Pi_{Stable}}$. Therefore, by referring to Lemma 3.3 again, one can conclude the existence of a deterministic stationary policy, say φ_2, such that

$$v_{max} = \sup_{\pi \in \Pi_{Stable}} V(\gamma, \pi, c_0) = - \inf_{\pi \in \Pi_{Stable}} V(\gamma, \pi, -c_0) = -V(\gamma, \varphi_2, -c_0) = V(\gamma, \varphi_2, c_0).$$

Thus, the extreme points of $\mathcal{V}$ are generated by deterministic stationary policies for the case of $N = 0$.

Suppose the statement holds for the case of $N = k - 1$, i.e., suppose for any CTMDP model with $N = k - 1$ constraints satisfying the corresponding Conditions 2.1, 3.1, 3.2, and 3.3 in case S is denumerable, and additionally Condition 3.4 in case S is uncountable, it holds that each extreme point v^{ex} of $\mathcal{V}$ is generated by a deterministic stationary policy.

Now consider the case of $N = k$, i.e., consider a CTMDP with k constraints satisfying Conditions 2.1, 3.1, 3.2, and 3.3 in case S is denumerable, and additionally Condition 3.4 in case S is uncountable.

It follows from its definition that the extreme point $v^{ex} = (v_0^{ex}, v_1^{ex}, \ldots, v_k^{ex})$ is not in the interior of $\mathcal{V} \subseteq \mathbb{R}^{k+1}$. So by the supporting hyperplane theorem [5], there exists a hyperplane

$$\mathcal{H} = \left\{ x = (x_0, x_1, \ldots, x_k) \in \mathbb{R}^{k+1} : \sum_{i=0}^{k} \lambda_i' x_i = \rho \right\}, \tag{15}$$

where $\lambda_i' \in \mathbb{R}, i = 0, 1, \ldots, k$, which are not all equal to zero, and $\rho \in \mathbb{R}$ are fixed constants defining the underlying hyperplane, such that

$$\sum_{i=0}^{k} \lambda_i' v_i^{ex} = \rho \leq \sum_{i=0}^{k} \lambda_i' v_i \ \forall \ v = (v_0, v_1, \ldots, v_k) \in \mathcal{V}.$$

Here, we take $\lambda_k' \neq 0$ without loss of generality, for otherwise one only needs re-order the cost rates. Note that the above equality and inequality can be equivalently written as

$$V\left(\gamma, \pi^{ex}, \sum_{i=0}^{k} \lambda_i' c_i\right) = \rho \leq V\left(\gamma, \pi, \sum_{i=0}^{k} \lambda_i' c_i\right) \ \forall \ \pi \in \Pi_{Stable}, \tag{16}$$

where π^{ex} is a stable policy that generates v^{ex}. In other words, π^{ex} is an optimal policy to the unconstrained CTMDP problem $V(\gamma, \pi, \sum_{i=0}^{k} \lambda_i' c_i) \to \min_{\pi \in \Pi_{Stable}}$, and so

$$\rho = v^*(\vec{\lambda'}), \tag{17}$$

where $\vec{\lambda'} := (\lambda_0', \ldots, \lambda_k')$, and $v^*(\vec{\lambda'})$ is as in Lemma 3.3.

Let us define the set

$$\tilde{\mathcal{U}} := \mathcal{H} \bigcap \mathcal{V}, \tag{18}$$

which is nonempty, convex and compact because so are both $\mathcal{H}$ and $\mathcal{V}$. Moreover, $v^{ex} \in \tilde{\mathcal{U}}$ is also an extreme point of $\tilde{\mathcal{U}}$ because $\tilde{\mathcal{U}} \subseteq \mathcal{V}$. Below we construct an appropriate auxiliary CTMDP model (see (20)), whose space of relevant performance vectors is denoted by $\hat{\mathcal{V}}$, which will be proved to coincide with $\tilde{\mathcal{U}}$.

Recalling the definition of the set $B(x)$ as in Lemma 3.3, we now formally define, for each $x \in S$, in case S is denumerable

$$\hat{A}(x) := B(x);$$

and in case S is uncountable

$$\hat{A}(x) := \begin{cases} B(x) & \text{if } \ x \in S_{\pi^{ex}}^{\vec{\lambda'}}, \\ \{\psi^*(x)\} & \text{if } x \notin S_{\pi^{ex}}^{\vec{\lambda'}}, \end{cases}$$

where $S_{\pi^{ex}}^{\vec{\lambda'}}$ and ψ^* are from Lemma 3.3.

We have the following three observations.

Observation 1. For each $x \in S$, the corresponding set $\hat{A}(x) \subseteq A(x)$ is nonempty compact.

Indeed, we have $\hat{A}(x)$ is closed for any $x \in S$ because of the definition of $\hat{A}(x)$ and the fact that the function

$$H(x, a) := \sum_{i=0}^{N} \lambda_i' c_i(x, a) + \int_S u_2^*(y) q(dy|x, a) \tag{19}$$

is continuous on $A(x)$ for each $x \in S$ by the virtue of [24, Lem.8.3.7]. Now the compactness of $\hat{A}(x)$ follows from its closedness and the compactness of $A(x)$; see Remark 3.1.

The next two observations are obvious in case S is denumerable, so that we shall only justify them for the case of S being uncountable.

Observation 2. The set

$$\hat{\mathbb{K}} := \{(x,a) : x \in S, a \in \hat{A}(x)\} \subseteq \mathbb{K}$$

is in $\mathcal{B}(S \times A)$.

Indeed, for each closed subset $F \subseteq A$,

$$\begin{aligned}\left\{x \in S : \hat{A}(x) \bigcap F \neq \emptyset\right\} &= \left\{x \in S^{\vec{\lambda}'}_{\pi^{ex}} : \inf_{a \in A(x)\bigcap F} H(x,a) = v^*(\vec{\lambda}')\right\} \\ &\quad \bigcup \left\{x \in S \setminus S^{\vec{\lambda}'}_{\pi^{ex}} : \psi^*(x) \in F\right\}.\end{aligned}$$

Since $A(x) \bigcap F$ is compact, see Remark 3.1, and the function $H(x,a)$ defined by (19) is continuous in $a \in A(x)$ for any $x \in S^{\vec{\lambda}'}_{\pi^{ex}}$ as observed earlier, the function $\inf_{a \in A(x)\bigcap F} H(x,a)$ is measurable on $S^{\vec{\lambda}'}_{\pi^{ex}}$ (by [25, 26] and Proposition D.5 in [23]). This, together with the fact that $\{x \in S \setminus S^{\vec{\lambda}'}_{\pi^{ex}} : \psi^*(x) \in F\}$ is measurable, implies that the set $\{x \in S : \hat{A}(x) \bigcap F \neq \emptyset\}$ is a measurable subset of S, asserting that the multifunction $x \to \hat{A}(x)$ is measurable. It follows from this fact and Observation 1 that $\hat{\mathbb{K}}$ is a measurable subset of $S \times A$; see [26] or Proposition D.4 of [23].

Observation 3. The set $\hat{\mathbb{K}}$ contains the graph of a measurable mapping from S to A.

Indeed, by Observations 1 and 2 above, Proposition D.5 in [23] ensures the existence of a measurable mapping g from S to A such that $g(x) \in \hat{A}(x)$ for each $x \in S$.

Based on the above three observations, we legally have an auxiliary $\hat{A}$-CTMDP model

$$\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}, \tag{20}$$

where $q(dy|x,a)$ and $c_i(x,a)$ are understood as their corresponding restrictions on $\hat{\mathbb{K}} \subseteq \mathbb{K}$. It is also an immediate consequence of those observations that the corresponding versions of Conditions 2.1, 3.1, 3.2, and 3.3 in case S is denumerable, and additionally Condition 3.4 in case S is uncountable, are satisfied by this auxiliary CTMDP model. In particular, Condition 3.1(c) is satisfied by an increasing sequence of compact (in the topology relative to $\hat{\mathbb{K}}$) sets $\hat{K}_m := \hat{\mathbb{K}} \bigcap K_m \uparrow \hat{\mathbb{K}}$, where $K_m \uparrow \mathbb{K}$ are the compact sets coming from Condition 3.1(c) for the original model. Indeed, this follows from the fact that

$$\begin{aligned}&\lim_{m\to\infty} \inf_{(x,a)\in\hat{\mathbb{K}}\setminus\hat{K}_m} \frac{w(x)}{w'(x)} = \lim_{m\to\infty} \inf_{(x,a)\in\hat{\mathbb{K}}\setminus K_m} \frac{w(x)}{w'(x)} \\ \geq\ &\lim_{m\to\infty} \inf_{(x,a)\in\mathbb{K}\setminus K_m} \frac{w(x)}{w'(x)} = \infty,\end{aligned}$$

where the last equality is due to the fact that Condition 3.1(c) is satisfied by the original model. It is worthwhile to mention that any policy in the auxiliary $\hat{A}$-CTMDP model is also one in the original average CTMDP model $\{S, A, A(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}$.

We claim that for this auxiliary $\hat{A}$-CTMDP model, the space of relevant performance vectors $\hat{\mathcal{V}}$ is the same as $\tilde{\mathcal{U}}$ defined by (18). To see this, we firstly show

$$\hat{\mathcal{V}} \subseteq \tilde{\mathcal{U}}.$$

Indeed, any stable policy for the auxiliary model $\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}$ is also a stable policy for the original CTMDP model $\{S, A, A(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}$, implying that any point in $\hat{\mathcal{V}}$ is also in $\mathcal{V}$. On the other hand, under the conditions of this theorem (especially Condition 3.4(b) in case S is uncountable), by the definition of $\hat{A}(x)$ and Lemma 3.3 (see also (27) in its proof), we have that any point in $\hat{\mathcal{V}}$ is also in the hyperplane $\mathcal{H}$ defined by (15). It follows from these facts and (18) that $\hat{\mathcal{V}} \subseteq \tilde{\mathcal{U}}$.

Secondly, we show

$$\tilde{\mathcal{U}} \subseteq \hat{\mathcal{V}}.$$

To this end, consider an arbitrary point $(V(\gamma,\pi,c_0),\ldots,V(\gamma,\pi,c_k)) \in \tilde{\mathcal{U}}$, where π is a stable policy for the original CTMDP model $\{S, A, A(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}$ and

$$V\left(\gamma,\pi,\sum_{i=0}^{k}\lambda_i' c_i\right) = v^*(\vec{\lambda'}). \tag{21}$$

In what follows, we show that

$$(V(\gamma,\pi,c_0),\ldots,V(\gamma,\pi,c_k)) = (V(\gamma,\hat{\pi},c_0),\ldots,V(\gamma,\hat{\pi},c_k)),$$

where $\hat{\pi}$ is a stable policy for the model $\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^k, \gamma\}$.

In case S is denumerable, we define $\hat{\pi}$ by

$$\hat{\pi}(da|x) = \pi(da|x),\ \forall\ x \in S_\pi^{\vec{\lambda'}}$$

and

$$\hat{\pi}(da|x) = I\{\psi^*(x) \in da\},\ \forall\ x \in S \setminus S_\pi^{\vec{\lambda'}},$$

where the set $S_\pi^{\vec{\lambda'}}$ is defined as in Lemma 3.3.

Now consider the case of S being uncountable. By Lemma 3.3(c) and (21) we have

$$\eta^\pi(S \setminus S_\pi^{\vec{\lambda'}}, A) = 0,\ \eta^{\pi^{ex}}(S \setminus S_{\pi^{ex}}^{\vec{\lambda'}}, A) = 0, \tag{22}$$

which, together with Condition 3.4(b), imply that

$$\eta^\pi(S_\pi^{\vec{\lambda'}} \bigcap S_{\pi^{ex}}^{\vec{\lambda'}}, A) = 1. \tag{23}$$

Here we recall that the sets $S_\pi^{\vec{\lambda'}}$, $S_{\pi^{ex}}^{\vec{\lambda'}}$ are defined as in Lemma 3.3(c). So by Lemma 3.3 and the definition of $\hat{A}(x)$, $\pi(da|x)$ is concentrated on $\hat{A}(x)$ for all $x \in S_\pi^{\vec{\lambda'}} \bigcap S_{\pi^{ex}}^{\vec{\lambda'}}$. We define a policy $\hat{\pi}$ such that

$$\hat{\pi}(da|x) = \pi(da|x),\ \forall\ x \in S_\pi^{\vec{\lambda'}} \bigcap S_{\pi^{ex}}^{\vec{\lambda'}}$$

and

$$\hat{\pi}(da|x) = I\{\psi^*(x) \in da\},\ \forall\ x \in S \setminus (S_\pi^{\vec{\lambda'}} \bigcap S_{\pi^{ex}}^{\vec{\lambda'}}).$$

In either of the above two cases, $\hat{\pi}$ is a stable policy for the auxiliary CTMDP model because (by (22)-(23) in case S is uncountable)

$$\eta^\pi(dx,da) = \eta^\pi(dx,A)\pi(da|x) = \eta^\pi(dx,A)\hat{\pi}(da|x).$$

It follows from the last equalities that

$$(V(\gamma,\pi,c_0),\dots,V(\gamma,\pi,c_k)) = (V(\gamma,\hat{\pi},c_0),\dots,V(\gamma,\hat{\pi},c_k)) \in \hat{\mathcal{V}}.$$

Consequently, $\tilde{\mathcal{U}} \subseteq \hat{\mathcal{V}}$ because the point $(V(\gamma,\pi,c_0),\dots,V(\gamma,\pi,c_k)) \in \tilde{\mathcal{U}}$ is arbitrarily fixed.

Therefore, $\hat{\mathcal{V}} = \tilde{\mathcal{U}}$, i.e., the auxiliary $\hat{A}$-CTMDP model

$$\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^{k}, \gamma\}$$

has the space of relevant performance vectors the same as the space $\tilde{\mathcal{U}}$ defined by (18), as claimed above. Below we legally study the space $\tilde{\mathcal{U}}$ as the space of relevant performance vectors for the auxiliary $\hat{A}$-CTMDP model $\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^{k}, \gamma\}$, and since the fixed extreme point v^{ex} of $\mathcal{V}$ is also an extreme point of $\tilde{\mathcal{U}} = \hat{\mathcal{V}}$, and any deterministic stationary policy for the auxiliary $\hat{A}$-CTMDP model is also one for the original CTMDP model, to complete the inductive argument, our objective becomes to show that v^{ex} is generated by a deterministic stationary policy for the auxiliary $\hat{A}$-CTMDP model $\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^{k}, \gamma\}$.

For the auxiliary model, a deterministic stationary policy generates the point $v^{ex} = (v_0^{ex}, v_1^{ex}, \dots, v_k^{ex})$ if and only if it generates $(v_0^{ex}, v_1^{ex}, \dots, v_{k-1}^{ex})$ because

$$v_k^{ex} = \frac{v^*(\vec{\lambda}') - \sum_{i=0}^{k-1} \lambda_i' v_i^{ex}}{\lambda_k'}, \tag{24}$$

see (15)-(18); recall that $\hat{\mathcal{V}} = \tilde{\mathcal{U}}$ and $\lambda_k' \neq 0$. So it is equivalent to consider the auxiliary CTMDP model

$$\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^{k-1}, \gamma\}$$

with only $k-1$ constraints, for which we denote the space of relevant performance vectors by $\hat{\mathcal{V}}' \subseteq \mathbb{R}^k$.

For this CTMDP model with $k-1$ constraints, the corresponding versions of Conditions 2.1, 3.1, 3.2, and 3.3 in case S is denumerable, and additionally Condition 3.4 in case S is uncountable, are all satisfied, because so are they by the auxiliary model $\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^{k}, \gamma\}$ with k constraints. Since $(v_0^{ex}, v_1^{ex}, \dots, v_k^{ex})$ is an extreme point of $\tilde{\mathcal{U}} = \hat{\mathcal{V}}$, $(v_0^{ex}, v_1^{ex}, \dots, v_{k-1}^{ex})$ is an extreme point of $\hat{\mathcal{V}}'$, see (24). Therefore, by the inductive supposition, the extreme point $(v_0^{ex}, v_1^{ex}, \dots, v_{k-1}^{ex})$ is generated by a deterministic stationary policy φ for the CTMDP model

$$\{S, A, \hat{A}(x), q(dy|x,a), (c_i(x,a), d_i)_{i=0}^{k-1}, \gamma\}$$

with $k-1$ constraints and thus also for the original CTMDP model. It follows from this and (24) that the originally arbitrarily fixed extreme point $v^{ex} = (v_0^{ex}, v_1^{ex}, \dots, v_k^{ex})$ of $\mathcal{V}$ is generated by a deterministic stationary policy φ for the original CTMDP model. This completes the inductive argument, and the statement is thus proved. □

Appendix

Proof of Proposition 3.1. Note that $\mathcal{D}$ is a subset of $\mathcal{P}_{w'}(\mathbb{K})$ by Remark 3.2. The non-emptiness of $\mathcal{D}$ follows from the proof of [22, Thm.3.9(a)], and the convexity of $\mathcal{D}$ is evident, following from the definition of stable measures. Below we prove the compactness of $\mathcal{D}$.

Firstly, we prove that $\mathcal{D}$ is precompact in $(\mathcal{P}_{w'}(\mathbb{K}), \tau(\mathcal{P}_{w'}(\mathbb{K})))$. Since the function $w'(\cdot)$ is continuous, by Lemma 3.1, it is equivalent to proving the set $\tilde{\mathcal{D}} := T_{w'}(\mathcal{D})$, where $T_{w'}$ is defined by (4),

to be precompact in $(\mathcal{P}_1(\mathbb{K}), \tau(\mathcal{P}_1(\mathbb{K})))$, where we recall that the usual weak topology $\tau(\mathcal{P}_1(\mathbb{K}))$ is metrizable. Then for any $\tilde{\eta} = T_{w'}(\eta) \in \tilde{\mathcal{D}}$, where $\eta \in \mathcal{D}$, it holds that

$$\int_{\mathbb{K}} \frac{w(x)}{w'(x)} \tilde{\eta}(dx, da) = \frac{\int_{\mathbb{K}} w(x)\eta(dx, da)}{\int_{\mathbb{K}} w'(x)\eta(dx, da)} \leq \int_{\mathbb{K}} w(x)\eta(dx, da) \leq 1 + \frac{b}{\rho}, \tag{25}$$

where the first equality is by the definition of the mapping $T_{w'}$, the first inequality is by that $w'(x) \geq 1$, and the second inequality follows from that $\eta \in \mathcal{D}$ and the definition of stable measures, see (2). Since under Condition 3.1(c), the function $\frac{w}{w'}$ is a moment by [23, Def.E.7], it then follows from (25) that the family $\tilde{\mathcal{D}}$ is tight. Hence, one can refer to Prokhorov's theorem for the precompactness of $\tilde{\mathcal{D}}$. Thus, $\mathcal{D}$ is precompact in $(\mathcal{P}_{w'}(\mathbb{K}), \tau(\mathcal{P}_{w'}(\mathbb{K})))$.

Secondly, we show that $\mathcal{D}$ is w'-closed in $\mathcal{P}_{w'}(\mathbb{K})$. By Lemma 3.1, it suffices to consider the convergence of sequences. So let $\{\eta_n\}$ be a sequence in $\mathcal{D}$ such that $\eta_n \xrightarrow{w'} \overline{\eta}$, where $\overline{\eta} \in \mathcal{P}_{w'}(\mathbb{K})$. (Here η_n should not be confused with the empirical measures defined by (6).) Then on the one hand,

$$\begin{aligned}
\int_{\mathbb{K}} w(x)\overline{\eta}(dx, da) &= \lim_{m\uparrow\infty} \int_{\mathbb{K}} \min\{w(x), m\}\overline{\eta}(dx, da) \\
&= \lim_{m\uparrow\infty} \left(\lim_{n\to\infty} \int_{\mathbb{K}} \min\{w(x), m\}\eta_n(dx, da) \right) \\
&\leq \lim_{m\uparrow\infty} \overline{\lim_{n\to\infty}} \int_{\mathbb{K}} w(x)\eta_n(dx, da) \leq \lim_{m\to\infty} \left(1 + \frac{b}{\rho}\right) = 1 + \frac{b}{\rho},
\end{aligned}$$

where the first equality is by Levy's monotone convergence theorem, the second equality is by the continuity of w and the convergence of $\{\eta_n\}$, and the second inequality is by that $\eta_n \in \mathcal{D}$ and the definition of stable measures, see (2). Hence, (2) is satisfied by the measure $\overline{\eta}$. On the other hand, if we consider the signed measure defined by $\int_{\mathbb{K}} q(dy|x, a)\overline{\eta}(dx, da)$, which is finite, then for any bounded continuous function $g(\cdot)$ on S, it holds that

$$\begin{aligned}
\int_S g(y) \int_{\mathbb{K}} q(dy|x, a)\overline{\eta}(dx, da) &= \int_{\mathbb{K}} \left(\int_S g(y)q(dy|x, a) \right) \overline{\eta}(dx, da) \\
&= \lim_{n\to\infty} \int_{\mathbb{K}} \int_S g(y)q(dy|x, a)\eta_n(dx, da) \\
&= \lim_{n\to\infty} \int_S g(y) \int_{\mathbb{K}} q(dy|x, a)\eta_n(dx, da) = 0,
\end{aligned}$$

where the second equality is by that $\int_S g(y)q(dy|x, a)$ is continuous and w'-bounded on $\mathbb{K}$, and $\eta_n \xrightarrow{w'} \overline{\eta}$, and the last equality is by (3). This, by [45, Lem. 2.3], implies that the signed measure $\int_{\mathbb{K}} q(dy|x, a)\overline{\eta}(dx, da)$ is equal to zero, and (3) is satisfied by the measure $\overline{\eta}$. Thus, both conditions of Definition 3.1 are satisfied by $\overline{\eta}$, i.e., $\overline{\eta} \in \mathcal{D}$. Consequently, $\mathcal{D}$ is w'-closed in $\mathcal{P}_{w'}(\mathbb{K})$.

Finally, it follows from the closedness and precompactness that $\mathcal{D}$ is w'-compact in $\mathcal{P}_{w'}(\mathbb{K})$. □

Proof of Lemma 3.2. (a) As in the proof of [36, Lem.A3], for each $i = 0, 1, \ldots, N$, there exists an increasing sequence of w'-bounded continuous functions $c_{i,m}(x, a)$ on $\mathbb{K}$ and a constant $\overline{c} \in \mathbb{R}$ such that $c_{i,m}(x, a) \uparrow c_i(x, a)$ and $\sup_{i,m} |c_{i,m}(x, a)| \leq \overline{c}w'(x)$ for all $(x, a) \in \mathbb{K}$. It follows from this and a standard argument that for any convergent sequence $\eta_n \xrightarrow{w'} \eta$, where $\eta_n, \eta \in \mathcal{P}_{w'}(\mathbb{K})$,

$$\underline{\lim}_{n\to\infty} \int_{\mathbb{K}} c_i(x, a)\eta_n(dx, da) \geq \int_{\mathbb{K}} c_i(x, a)\eta(dx, da),$$

i.e., $\int_{\mathbb{K}} c_i(x, a)\eta(dx, da)$ is lower semicontinuous in $\eta \in \mathcal{P}_{w'}(\mathbb{K})$.

(b) This part of the lemma was presented as Theorem 3.9(a) in [22] assuming the continuity of the cost rates $c_i(x,a), i = 0, 1, \ldots, N$. The proof of [22, Thm.3.9(a)] still applies to lower semicontinuous functions $c_i(x,a), i = 0, 1, \ldots, N$; one only needs legally change the first equality in [22, Eqn.(3.11)] to inequality "$\geq$" by using the fact in (a). □

Proof of Proposition 3.2. Under the imposed conditions, for any $i = 0, 1, \ldots, N$, it follows from Lemma 3.2 (a) that the function $\int_{\mathbb{K}} c_i(x,a)\eta(dx,da)$ is lower semicontinuous in $\eta \in \mathcal{D}$. Therefore, the space of feasible stable measures defined by $\mathcal{D}^F := \{\eta \in \mathcal{D} : \int_{\mathbb{K}} c_j(x,a)\eta(dx,da) \leq d_j, j = 1, 2, \ldots, N\}$ is w'-closed in $\mathcal{D}$. Since $\mathcal{D}$ is w'-compact by Proposition 3.1, $\mathcal{D}^F$, as a closed subset of $\mathcal{D}$, is w'-compact, too. The statement now follows. □

Proof of Lemma 3.3. Part (a) follows from [19, Thm.4.2] for this statement.

For part (b), it follows from part (a) that for each $(x,a) \in \mathbb{K}$

$$c^{\vec{\lambda}}(x,a) + \int_S u_2^*(y)q(dy|x,a) \geq v^*(\vec{\lambda}) \geq c^{\vec{\lambda}}(x,\varphi^*(x)) + \int_S u_1^*(y)q(dy|x,\varphi^*(x)). \tag{26}$$

For any stable policy π, let $\eta^\pi(dx,da)$ denote the stable measure corresponding to π. Then from the first inequality in (26) one obtains

$$\begin{aligned} v^*(\vec{\lambda}) &\leq \int_{\mathbb{K}} c^{\vec{\lambda}}(x,a)\pi(da|x)\eta^\pi(dx,A) + \int_S \eta^\pi(dx,A)\int_S u_2^*(y)\int_{A(x)} q(dy|x,a)\pi(da|x) \\ &= \int_{\mathbb{K}} c^{\vec{\lambda}}(x,a)\pi(da|x)\eta^\pi(dx,A) + \int_S u_2^*(y)\int_{\mathbb{K}} q(dy|x,a)\pi(da|x)\eta^\pi(dx,A) \\ &= V(\gamma,\pi,c^{\vec{\lambda}}), \end{aligned} \tag{27}$$

where the last equality is by Condition 3.4 and the definition of a stable policy, and the interchanges of the order of integration are all legal due to Fubini's theorem.

Similarly, from the second inequality in (26) we have $v^*(\vec{\lambda}) \geq V(\gamma,\varphi^*,c^{\vec{\lambda}})$, which, together with (27) and Condition 3.4, implies the statement of (b).

Part (c) follows from part (a) and (27).

Part (d) can be seen by applying the reasoning presented in Theorem 7.8 in Chapter 7 of [20] (the conditions assumed therein can be relaxed to the present setup). □

Acknowledgments.

We are grateful to the referees and the associate editor for the constructive comments and insightful remarks. Especially one referee brought to our attention the relevant statements of [7] and the Ph.D thesis of Junli Zheng [47]. Also Professor Mufa Chen (Beijing Normal University) kindly sent us his preprint [9]. This research is partially supported by NSFC and GDUPS.

References

[1] Aliprantis, C., Border, K. 2007. *Infinite Dimensional Analysis*. Springer, New York.

[2] Altman, E., Shwartz, A. 1991. Markov decision problems and state-action frequencies. *SIAM J. Control Optim.* **29**: 786-809.

[3] Altman, E. 1999. *Constrained Markov Decision Processes*. Chaptman &Hall/CRC, Boca Raton.

[4] Anderson, W. 1991. *Continuous-time Markov Chains*. Springer, New York.

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

[6] Borkar, V. 1994. Ergodic control of Markov chains with constraints–the general case. *SIAM J. Control Optim.* **32**: 176-186.

[7] Chen, M. 1986. Coupling for jump processes. *Acta Math. Sin. English Series* **2**: 123-136.

[8] Chen, M. 2004. *From Markov Chains to Non-equilibrium Particle Systems*. World Scientific, Singapore.

[9] Chen, M. 2015. Practical criterion for uniqueness of q-processes. *Chinese J. Appl. Probab. Stat.* **31**: 213-224.

[10] Dekker, R., Hordijk, A., Spieksma, F. 1994. On the relation between recurrence and ergodicity properties in denumrable Markov decision chains. *Math. Oper. Res.* **19**: 539-559.

[11] Dubins, L. 1962. On extreme points of convex sets. *J. Math. Anal. Appl.* **5**: 237-244.

[12] Feinberg, E., Shwartz, A. 1996. Constrained discounted dynamic programming. *Math. Oper. Res.* **21**: 922-944.

[13] Feinberg, E., Piunovskiy, A. 2000. Multiple objective nonatomic Markov decision processes with total reward criteria. *J. Math. Anal. Appl.* **247**: 45-66.

[14] Feinberg, E. 2002. Optimal control of average reward constrained continuous-time finite Markvo decision processes. *Proc. 41st IEEE CDC*, 3805-3810.

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

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

[17] Feinberg, E.. Reduction of discounted continuous-time MDPs with unbounded jump and reward rates to discrete-time total-reward MDPs. *Optimization, Control, and Applications of Stochastic Systems*, Hernandez-Hernandez, D., Minjarez-Sosa, A., eds. Springer, New York, 77-97.

[18] Glynn, P. 1994. Some topics in regenerative steady-state simulation. *Acta Appl. Math.* **34**: 225–236.

[19] Guo, X., Rieder, U. 2006. Average optimality for continuous-time Markov decision processes in Polish spaces. *Ann. Appl. Probab.* **16**: 730-756.

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

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

[22] Guo, X., Huang, Y., Song, X. 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.

[23] Hernández-Lerma, O., Lasserre, J. 1996. *Discrete-Time Markov Control Processes*. Springer, New York.

[24] Hernández-Lerma, O., Lasserre, J. 1999. *Further Topics in Discrete-Time Markov Control Processes*. Springer, New York.

[25] Himmelberg, C. 1975. Measurable relations. *Fund. Math.* **87**: 53-72.

[26] Himmelberg, C., Parthasarathy, T., Van Vleck, F. 1976. Optimal plans for dynamic programming problems. *Math. Oper. Res.* **1**: 390-394.

[27] Hordjik, A., Yushkevich, A. 1999. Blackwell optimality in the class of all policies in Markov decision chains with a Borel state space and unbounded rewards. *Math. Meth. Oper. Res.* **50**: 421-448.

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

[29] Kitaev, M. 1986. Semi-Markov and jump Markov controlled models: average cost criterion. *Theory. Probab. Appl.* **30**: 272-288.

[30] Kitaev, M., Rykov, V. 1995. *Controlled Queueing Systems*. CRC Press, Boca Raton.

[31] Kumar, P., Meyn, S. 1996. Duality and linear programs for stability and performance analysis of queueing networks and scheduling policies. *IEEE Trans. Automat. Control* **41**: 4-17.

[32] Meyn, S., Tweedie, R. 1993. Stability of Markov processes II: continuous-time processes and sampled chains. *Adv. Appl. Probab.* **25**: 487-517.

[33] Meyn, S., Tweedie, R. 1993. Stability of Markov processes III: Forster-Lyapunov criteria for continous time processes. *Adv. Appl. Probab.* **25**: 518-548.

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

[35] Piunovskiy, A. 2004. Bicriteria optimization of a queue with a controlled input stream. *Queueing. Syst.* **48**: 159-184.

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

[37] Prieto-Rumeau, T., Hernández-Lerma, O. 2008. Ergodic control of continuous-time Markov chains with pathwise constraints. *SIAM J. Control. Optim.* **47**: 1888-1908.

[38] Prieto-Rumeau, T., Hernández-Lerma, O. 2010. The vanishing discount approach to constrained continuous-time controlled Markov chains. *Syst. Control. Lett.* **59**: 504-509.

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

[40] Prieto-Rumean. T., Lorenzo, J. 2015. Approximation of zero-sum continuous-time Markov games under the discounted payoff criterion. *Top* **23**: 799-836.

[41] Prieto-Rumeau, T., Hernández-Lerma, O. 2012. Uniform ergodicity of continuous-time controlled Markov chains: a survey and new results. *Ann. Oper. Res.*, to appear.

[42] Serfozo, R. 1982. Convergence of Lebesgue integrals with varying measures. *Sankhya* **44**:280-402.

[43] Spieksma, F. 1990. *Geometrically Ergodic Markov Chains and the Optimal Control of Queues.* Ph.D. Thesis. University of Leiden, 1990.

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

[45] Varadarajan, V. 1958. Weak convergence of measures on separable metric spaces. *Sankhyā* **19**: 15-22.

[46] Ye, L., Guo, X. 2010. New sufficient conditions for average optimality in contiuous-time Markov decision processes. *Math. Meth. Oper. Res.* **72**: 75-94.

[47] Zheng, J. 1993. *Phase Transitions of Ising Model on Lattice Fractals, Martingale Approach for q-processes* (in Chinese). Ph.D. Thesis. Beijing Normal University.