Ann. of Appl. Math.
35:4(2019), 449-464

CONVERGENCE OF CONTROLLED MODELS FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES WITH CONSTRAINED AVERAGE CRITERIA*†

Wenzhao Zhang, Xianzhu Xiong‡

(*College of Math. and Computer Science, Fuzhou University, Fuzhou 350108, Fujian, PR China*)

Abstract

This paper attempts to study the convergence of optimal values and optimal policies of continuous-time Markov decision processes (CTMDP for short) under the constrained average criteria. For a given original model $\mathcal{M}_\infty$ of CTMDP with denumerable states and a sequence $\{\mathcal{M}_n\}$ of CTMDP with finite states, we give a new convergence condition to ensure that the optimal values and optimal policies of $\{\mathcal{M}_n\}$ converge to the optimal value and optimal policy of $\mathcal{M}_\infty$ as the state space S_n of $\mathcal{M}_n$ converges to the state space S_∞ of $\mathcal{M}_\infty$, respectively. The transition rates and cost/reward functions of $\mathcal{M}_\infty$ are allowed to be unbounded. Our approach can be viewed as a combination method of linear program and Lagrange multipliers.

Keywords continuous-time Markov decision processes; optimal value; optimal policies; constrained average criteria; occupation measures

2000 Mathematics Subject Classification 90C40; 60J27

1 Introduction

Markov decision processes have wide application in queueing system, telecommunications systems, etc.; see, for instance, [2, 11, 13, 16, 18] and the reference therein. The existence and computation of optimal value and optimal policies form a hot research area in Markov decision processes. The basic method to study the existence of optimal policies include the dynamic programming approach, the linear programming and duality programming method. Based on above methods, the value iteration algorithms, policy iteration algorithms, linear programming algorithms for unconstrained optimality problems and linear programming algorithms for constrained optimality problems have been proposed; see, for instance, [4, 11, 13, 15]. However,

*The first author is supported by the NNSF of China (No.11801080).
†Manuscript received June 3, 2019; Revised October 10, 2019
‡Corresponding author. E-mail: xiongxianzhu2001@sina.com

these algorithms are only adapt to tackle the optimality problems with finite states. It is natural to use finite-state models to approximate the original model with denumerable state space or general Borel state space. Hence, from the theoretical and practical point of view, the convergence of optimal values and optimal policies are important and interesting issues in Markov decision processes.

In the discrete-time context, [3] considered the convergence of optimal value and optimal policies of Markov decision processes with denumerable states under the constrained expected discounted cost criteria. [5, 6] developed the approximation method of optimal value and optimal policies of Markov decision processes with Borel state and action spaces under the constrained expected discounted cost criteria.

In the continuous-time formulation, [16] studied the convergence of optimal value and optimal policies of Markov decision processes with denumerable states under the expected discounted cost and average cost criteria. [17, 18] developed an approximation procedure for CTMDP with denumerable state space under the finite-horizon expected total cost criterion and risk-sensitive finite-horizon cost criterion, respectively. For constrained optimal problem, [12] proposed an approach based on occupation measures to study the convergence problem of optimal value and optimal policies, and gave condition imposed on the original model with denumerable states to ensure the original model can be approximated by a sequence of CTMDP with finite states.

In this paper, we consider the similar convergence problem as in [12] with denumerable states but under the constrained expected average criteria. More precisely, the original controlled model has the following features: 1) The state space is denumerable and the action space is a Polish space; 2) the transition rates, cost and reward functions may be unbounded from above and from below. Firstly, by introducing the average occupation measures and Lagrange multipliers, we prove that the constrained optimality problem of each model $\mathcal{M}_n$ of CTMDP equals to a unconstrained optimality problem, and deduce the optimality equation which includes some Lagrange multipliers. These results are extension of the results in [16] for constrained optimality problem with one constraint. Then, we derive the bound of the Lagrange multipliers in each model $\mathcal{M}_n$. Secondly, according to the optimality equations, we give the exact bound of of the optimal values between the finite-state model $\mathcal{M}_n$ and the original model $\mathcal{M}_\infty$. Finally, using some approximation properties of expected average reward/cost, we obtain the asymptotic convergence of optimal policies of finite-state models to the optimal policy of the original model.

The rest of the paper is organized as follows. In Section 2, we introduce the constrained average model we are concerned with. In Section 3, we deduce the optimality equation of each constrained model $\mathcal{M}_n$ and give the error bounds of the

Lagrange multipliers in optimality equations. In Section 4, we obtain the asymptotic convergence of optimal values of finite-state models to the optimal value of the original model and asymptotic convergence of policies of finite-state models to the optimal policy of the original model.

2 The Models

In this section we introduce the models we are concerned with.

Notation If X is a Polish space, we denote by $\mathcal{B}(X)$ its Borel σ-algebra, by $\mathcal{P}(X)$ the set of probability measures on $\mathcal{B}(X)$ endowed with the topology of weak convergence, by $\mathbb{B}_b(X)$ the Banach space of all bounded measurable functions on X, by $\mathbb{C}_b(X)$ the Banach space of all bounded continuous functions on X. Let $\mathbb{N} := \{1, 2, \cdots\}$, $\overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$, $\mathbb{R}_+ := (0, \infty)$ and $\mathbb{R}_+^0 := [0, \infty)$.

Consider the sequence of models $\{\mathcal{M}_n\}$ for constrained CTMDP:

$$\mathcal{M}_n := \big\{S_n, (A(i) \subseteq A, i \in S_n), q_n(\cdot|i,a), r_n(i,a), (c_n^l(i,a), d_n^l, 1 \le l \le p), \gamma_n\big\}, \quad n \in \overline{\mathbb{N}}, \tag{2.1}$$

where S_n is the *state space*. We assume $S_n := \{0, 1, \cdots, n\}$ for each $n \in \mathbb{N}$ and $S_\infty := \{0, 1, \cdots\}$. As a consequence, for each $i \in S_\infty$, we can define $n(i) := \min\{n \ge 1, i \in S_n\}$. The set A is the *action space* which is assumed to be a Polish space and the set $A(i) \in \mathcal{B}(A)$ represents the set of all available actions or decisions at state $i \in S_n$ for each $n \in \overline{\mathbb{N}}$. Let $K_n := \big\{(i,a) | i \in S_n, a \in A(i)\big\}$ represent the set of all feasible state-action pairs. For fixed $n \in \overline{\mathbb{N}}$, the function $q_n(\cdot|i,a)$ in (2.1) denotes the transition rates, that is, $q_n(j|i,a) \ge 0$ for all $(i,a) \in K_n$ and $i \ne j$. Furthermore, $q_n(i|j,a)$ is assumed to be *conservative*, that is

$$\sum_{j \in S_n} q_n(j|i,a) = 0, \quad \text{for all } (i,a) \in K_n, \tag{2.2}$$

and *stable*, that is

$$q_n^*(i) := \sup_{a \in A(i)} |q_n(i|i,a)| < \infty, \quad \text{for each } i \in S_n. \tag{2.3}$$

Moreover, $q_n(j|i,a)$ is a measurable function on $A(i)$ for each fixed $i, j \in S_n$. Finally, r_n corresponds to the reward function that is to be maximized, and c_n^l corresponds to the cost function on which the constraint $d_n^l \in \mathbb{R}$ is imposed for each $1 \le l \le p$. The γ_n denotes the initial distribution for $\mathcal{M}_n$.

Next, we briefly recall the construction of the stochastic basis $(\Omega_n, \mathcal{F}_n, \{\mathcal{F}_{t,n}\}_{t \ge 0}, P_{\gamma_n,n}^\pi)$ for each $n \in \mathbb{N}$. Let $i_\infty \notin S_\infty$ be an isolated point and $i_\infty \notin S_\infty$, $S_n^* := S_n \cup \{i_\infty\}$, $\Omega_n := (S_n \times (\mathbb{R}_+ \times S_n)^\infty) \cup \bigcup_{m=0}^{\infty} \big(S_n \times (\mathbb{R}_+ \times S_n)^m \times (\{\infty\} \times \{i_\infty\})^\infty\big)$. Thus, we obtain the sample space $(\Omega_n, \mathcal{F}_n)$, where $\mathcal{F}_n$ is the standard Borel σ-algebra. For

each $m \geq 1$ and each sample $\omega = (i_0, \theta_1, i_1, \cdots, \theta_{n-1}, i_{n-1}, \cdots) \in \Omega_n$, we define some maps on Ω_n as follows: $T_0(\omega) := 0$, $X_0(\omega) := i_0$, $\Theta_m(\omega) := \theta_m$, $T_m(\omega) := \sum_{n=1}^{m} \theta_n$, $T_\infty(\omega) := \lim_{n\to\infty} T_m(\omega)$, $X_m(\omega) := i_m$. Here, Θ_m, T_m, X_m denote the sojourn time, jump moment and the state of the process on the interval $[T_m, T_{m+1})$, respectively. Define a process $\{\xi_t, t \geq 0\}$ on $(\Omega_n, \mathcal{F}_n)$ by

$$\xi_t(\omega) := \sum_{m\geq 0} I_{\{T_m(\omega)\leq t<T_{m+1}(\omega)\}} i_m + I_{\{T_\infty(\omega)\leq t\}} i_\infty \quad \text{for each } \omega \in \Omega_n.$$

In what follows, $h_n(\omega) := (i_0, \theta_1, i_1, \cdots, \theta_n, i_n)$ is the n-component internal history, the argument $\omega = (i_0, \theta_1, i_1, \cdots, \theta_n, i_n, \cdots) \in \Omega_n$ is often omitted. Since we do not consider the process after T_∞, i_∞ is regarded as absorbing. Define $A(i_\infty) := a_\infty$, where $a_\infty \notin A$ is a isolated point, $A^* := A \cup \{a_\infty\}$ and $q(i_\infty|i_\infty, a_\infty) := 0$. Let $\mathcal{F}_{t,n} := \sigma(\{T_m \leq s, X_m = j\} : j \in S_n, s \leq t, m \geq 0)$ be the internal history to time t for the game model $\mathcal{G}$, $\mathcal{F}_{s-,n} := \bigvee_{t<s} \mathcal{F}_{t,n}$, $\mathcal{P}_n := \sigma(C \times \{0\}(C \in \mathcal{F}_0), C \times (s, \infty)(C \in \mathcal{F}_{s-,n}, s > 0))$ which denotes the predictable σ-algebra on $\Omega_n \times \mathbb{R}^0_+$.

Below we introduce the concept of policies. Let Φ_n denote the set all kernels on A^* given S^*_n.

Definition 2.1 (i) A $\mathcal{P}_n$-measurable transition probability function $\pi(\cdot|\omega, t)$ on $(A^*, \mathcal{B}(A^*))$, concentrated on $A(\xi_{t-}(\omega))$, is called a *randomized Markov policy* if there exists $\varphi(\cdot|\cdot, t) \in \Phi_n$ for each $t > 0$ such that $\pi(\cdot|\omega, t) = \varphi(\cdot|\xi_{t-}(\omega), t)$.

(ii) A randomized Markov policy π is said to be *randomized stationary* if there exists a stochastic kernel $\varphi \in \Phi_n$ such that $\pi(\cdot|\omega, t) = \varphi(\cdot|\xi_{t-}(\omega))$ for each $t > 0$. Such policies are denoted as φ.

We denote by Π_n the family of all randomized Makov policies of $\mathcal{M}_n$. The set of all stationary policies is denoted by Π^s_n. For each given $n \in \mathbb{N}$ and policy $\pi \in \Pi_n$, according to Theorem 4.27 in [14], there exists a unique probability measure $P^\pi_{\gamma_n,n}$ on $(\Omega_n, \mathcal{F}_n)$. Expectations with respect to $P^\pi_{\gamma_n,n}$ is denoted as $E^\pi_{\gamma_n,n}$. When $\gamma_n(i) = 1$, we write $P^\pi_{i,n}$ for $P^\pi_{\gamma_n,n}$ and $E^\pi_{i,n}$ for $E^\pi_{\gamma_n,n}$, respectively.

To guarantee the state processes $\{\xi_t, t \geq 0\}$ for each model $\mathcal{M}_n$ is nonexplosive, we impose the following so-called *drift conditions*.

Assumption 2.1 There exist a nondecreasing function $w \geq 1$ on S_∞, constants $\kappa_1 \geq \rho_1 > 0$, $L > 0$ and a finite set $C_n \subset S_n$ for each $n \in \overline{\mathbb{N}}$ such that

(a) $\lim_{i\to\infty} w(i) = \infty$;

(b) $\sum_{j\in S_n} q_n(j|i, a)w^2(j) \leq -\rho_1 w^2(i) + \kappa_1 I_{C_n}(i)$ for all $(i, a) \in K_n$, $n \in \overline{\mathbb{N}}$;

(c) $q^*_n(i) \leq Lw(i)$ for all $i \in S_n$, $n \in \overline{\mathbb{N}}$;

(d) $\sum_{i\in S_n} w^2(i)\gamma_n(i) < \infty$ for all $n \in \overline{\mathbb{N}}$.

Remark 2.1 Under Assumptions 2.1 (a)-(c), it follows from Theorem 2.13 in [16] that there exist constants $\rho'_\tau > 0$ and $\kappa'_\tau > 0$ for each $0 < \tau < 2$ such that

$$\sum_{j\in S_n} q_n(j|i,a)w^\tau(j) \le -\rho'_\tau w^\tau(i) + \kappa'_\tau I_{C_n}(i), \quad \text{for any } (i,a) \in K_n. \tag{2.4}$$

The expected average criteria $J_n^l(\gamma_n,\pi)$ for each given $n \in \mathbb{N}$ and $\pi \in \Pi_n$ are defined as follows:

$$J_n^0(\gamma_n,\pi) := \liminf_{T\to\infty} \frac{1}{T} E^\pi_{\gamma_n,n}\left[\int_0^T\int_A r_n(\xi_t,a)\pi(\mathrm{d}a|\xi_t,t)\mathrm{d}t\right], \tag{2.5}$$

$$J_n^l(\gamma_n,\pi) := \limsup_{T\to\infty} \frac{1}{T} E^\pi_{\gamma_n,n}\left[\int_0^T\int_A c_n^l(\xi_t,a)\pi(\mathrm{d}a|\xi_t,t)\mathrm{d}t\right], \quad \text{for all } 1\le l\le p. \tag{2.6}$$

Let $U_n := \{\pi \in \Pi_n |\, J_n^l(\gamma_n,\pi) \le d_n^l, 1 \le l \le p\}$, and $J_n^* = \sup_{\pi\in U_n} J_n^0(\gamma_n,\pi)$ be the set of all constrained policies and the optimal value of $\mathcal{M}_n(n \in \overline{\mathbb{N}})$, respectively.

Definition 2.2 (i) For any $n \in \overline{\mathbb{N}}$, a policy $\pi \in U_n$ is called an (constrained) optimal policy of $\mathcal{M}_n$ if $J_n^0(\gamma_n,\pi) = J_n^*$.

(ii) A sequence $\{\varphi_n\}$ with $\varphi_n \in \Pi_n^s$ for each $n \in \mathbb{N}$ is said to converge weakly to $\varphi \in \Pi_\infty^s$, if the sequence $\{\varphi_n(\cdot|i)\}$ converges weakly to $\varphi(\cdot|i)$ in $\mathcal{P}(A(i))$ for each $i \in S_\infty$ and $n \ge n(i)$. We denote it by $\varphi_n \to \varphi$.

3 Preliminary Results

For convenience, we define $c_n(i,a) := (c_n^1(i,a),\cdots,c_n^p(i,a))$ and $d_n := (d_n^1,\cdots,d_n^p)$ for each $n \in \overline{\mathbb{N}}$ and $(i,a) \in K_n$. Let e be the p-dimensional vector with all components equal to one. First, for the existence of an optimal policy π_n of $\mathcal{M}_n$, we introduce the following conditions from [9,16]:

Assumption 3.1 (a) For each $i \in S_n$, $A(i)$ is a compact set.

(b) The functions $q_n(j|i,\cdot)$, $r_n(i,\cdot)$ $c_n^l(i,\cdot)$ and $\sum_{j\in S_n} w(j)q_n(j|i,\cdot)$ are all continuous in $a \in A(i)$, for each fixed $n \in \overline{\mathbb{N}}$, $i,j \in S_n$ and $1 \le l \le p$.

(c) There exists a constant $M > 0$, such that $|r_n(i,a)| \le Mw(i)$ and $|c_n^l(i,a)| \le Mw(i)$ for all $n \in \overline{\mathbb{N}}$, $(i,a) \in K_n$ and $1 \le l \le p$.

(d) There exist constants $\eta > 0$ and $\kappa_2 \ge \rho_2 > 0$ such that $\sum_{j\in S_n} w^{2+\eta}(j)q_n(j|i,a) \le -\rho_2 w^{2+\eta}(i) + \kappa_2$ for each $n \in \overline{\mathbb{N}}$ and $(i,a) \in K_n$.

Remark 3.1 Assumption 3.1(a) implies that the space $\mathcal{P}(A(i))$ with the topology of weak convergence is also compact for each $i \in S_\infty$. Hence, by the Tychonoff's theorem, $\Pi_n^s = \prod_{i\in S_n} \mathcal{P}(A(i))$ is compact too.

Under Assumptions 2.1 and 3.1(c), we can obtain the finiteness of the expected average criteria.

Lemma 3.1[11,16] *Suppose that Assumptions* 2.1 *and* 3.1(c) *hold. Then*

$$|J_n^l(\gamma_n,\pi)| \le M\frac{\kappa_1'}{\rho_1'} \quad for\ each\ n \in \overline{\mathbb{N}},\ \pi \in \Pi_n\ and\ \ 0 \le l \le p,$$

where κ_1' *and* ρ_1' *are the constants in Remark* 2.1 *for* $\tau = 1$.

To obtain our main results, we consider the following assumptions:

Assumption 3.2 (a) $\lim\limits_{n\to\infty}\sup\limits_{a\in A(i)}|q_n(j|i,a)-q_\infty(j|i,a)|=0$, for each $i,j\in S_\infty$, where $n \ge \max\{n(i),n(j)\}$;

(b) $\lim\limits_{n\to\infty}\sup\limits_{a\in A(i)}|r_n(i,a)-r_\infty(i,a)|=0$ and $\lim\limits_{n\to\infty}\sup\limits_{a\in A(i)}|c_n^l(i,a)-c_\infty^l(i,a)|=0$ for each $i\in S_\infty$ and $1\le l\le p$, where $n\ge n(i)$;

(c) $\lim\limits_{n\to\infty} d_n^l = d_\infty^l$ for each $1\le l\le p$.

For each $n\in\overline{\mathbb{N}}$, a measurable function u on S_n (resp., K_n) is said to be with a finite w-norm if $||u||_w := \sup\limits_{i\in S_n}\frac{|u(i)|}{w(i)} < \infty$ (resp., $||u||_w := \sup\limits_{(i,a)\in K_n}\frac{|u(i,a)|}{w(i)} < \infty$). We denote by $\mathbb{B}_w(S_n)$ the Banach space of functions on S_n with finite w-norm and denote the set $\mathcal{C}_w(K_n) = \{u : K_n \to \mathbb{R}\big| u \text{ is continuous on } K_n \text{ and } \sup\limits_{(i,a)\in K_n}\frac{|u(i,a)|}{w(i)} < \infty\}$.

Assumption 3.3 For each $n\in\overline{\mathbb{N}}$ and $\varphi\in\Pi_n^s$, the corresponding Markov process ξ_t in each model $\mathcal{M}_n$ is irreducible.

Remark 3.2 (i) Under Assumptions 2.1and 3.3, Theorem 2.5 in [16] yields that for each $n\in\overline{\mathbb{N}}$ and $\varphi\in\Pi_n^s$, the Markov chain $\{\xi_t\}$ has a unique invariant probability measure, denoted by μ_n^φ.

(ii) Under Assumptions 2.1, 3.1 and 3.3, Theorem 2.11 in [16] implies that the control model (2.1) is uniformly w-exponentially ergodic, that is, there exist constants $\delta_n > 0$ and $\beta_n > 0$ such that

$$\sup_{\varphi\in\Pi_n^s}|E_{i,n}^\varphi\big(u(\xi_t)\big) - \mu_n^\varphi(u)| \le \beta_n \mathrm{e}^{-\delta_n t}\|u\|_w w(i),$$

for each $n\in\overline{\mathbb{N}}$, $u\in\mathbb{B}_w(S_n)$ and $t\ge 0$, where $\mu_n^\varphi(u) := \sum\limits_{j\in S_n} u(j)\mu_n^\varphi(j)$. Moreover, by Remark 2.1, we have $\mu_n^\varphi(w^\tau) := \sum\limits_{j\in S_n} w^\tau(j)\mu_n^\varphi(j) \le \frac{\kappa_\tau'}{\rho_\tau'}$ for each $0<\tau\le 2$, where $\rho_2' = \rho_1$ and $\kappa_2' = \kappa_1$.

(iii) Under Assumptions 2.1, 3.1 and 3.3, for each $n\in\overline{\mathbb{N}}$, $0\le l\le p$ and stationary policy $\varphi\in\Pi_n^s$, the $J_n^l(\gamma_n,\varphi)$ is a constant and does not depend on the initial state i, more precisely, $J_n^l(\gamma_n,\varphi) = \sum\limits_{j\in S_n} c_n^l(j,\varphi)\mu_n^\varphi(j) := g_n^l(\varphi)$.

Assumption 3.4(Slater condition) There exists a policy $\pi\in\Pi_\infty$ such that

$$J_\infty^l(\gamma_\infty,\pi) < d_\infty^l \quad \text{for all } 1\le l\le p. \tag{3.1}$$

For each $\varphi \in \Pi^s_\infty$, we denote by $\varphi|_{S_n}$ the restriction of φ in the set S_n as follows: $\varphi|_{S_n}(\cdot|i) := \varphi(\cdot|i)$ for each $i \in S_n$.

Lemma 3.2 *Suppose Assumptions* 2.1, 3.1, 3.2(a)-(b) *and* 3.3 *hold. Then, for each* $0 \le l \le p$,

$$\lim_{n\to\infty} \sup_{\varphi\in\Pi^s_\infty} |g^l_\infty(\varphi) - g^l_n(\varphi|_{S_n})| = 0.$$

This statement has been established by Theorem 4.21 in [16] for deterministic stationary policies. It is easy to extend the result to the class of randomized stationary policies.

For each $n \in \overline{\mathbb{N}}$ and $\varphi \in \Pi^s_n$, we define the measure $\widetilde{\mu}^\varphi_n$ by $\widetilde{\mu}^\varphi_n(i \times B) := \mu^\varphi_n(i)\varphi(B|i)$ for each $i \in S_n$ and $B \subseteq A(i)$. For measurable function w, let

$$\mathcal{P}_w(K_n) := \left\{\eta \in \mathcal{P}(K_n) | \int_{K_n} w(i)\eta(i, \mathrm{d}a) < \infty\right\}.$$

In particular, under Assumptions 2.1, 3.1 and 3.3, $\widetilde{\mu}^\varphi_n \in \mathcal{P}_w(K_n)$ for each $\varphi \in \Pi^s_n$ and $n \in \overline{\mathbb{N}}$. Now, we introduce the following sets

$$\Lambda_n := \{\widetilde{\mu}^\varphi_n | \, \varphi \in \Pi^s_n\},$$

and

$$\Lambda^f_n := \left\{\mu \in \Lambda_n | \int_{K_n} c^l_n(i, a)\mu(i, \mathrm{d}a) \le d^l_n, \text{ for each } 1 \le l \le p\right\} \quad \text{for each } n \in \overline{\mathbb{N}},$$

where the index "f" in Λ^f_n stands for "feasible".

Definition 3.1 The w-weak topology on $\mathcal{P}_w(K_n)$ is the coarsest topology for which all mappings

$$\mu \mapsto \int_{K_n} f\mathrm{d}\mu, \quad \text{where } f \in C_w(K_n)$$

are continuous.

The following lemma characters the w-weak topology; for a proof, see [8, Corollary A.45].

Lemma 3.3 *A sequence* $\{\mu_m\} \in \mathcal{P}_w(K_n)$ *converges* w*-weakly to* μ *if and only if* $\int_{K_n} f\mathrm{d}\mu_m \to \int_{K_n} f\mathrm{d}\mu$ *for every measurable function* f *which is* μ*-a.e continuous on* K_n *and for which exists a constant* c *such that* $|f| \le c \cdot w$ μ*-almost everywhere. In this case, we write* $\mu_m \xrightarrow{w} \mu$.

Lemma 3.4 *Suppose that Assumptions* 2.1, 3.1 *and* 3.3 *hold. Then, the set* Λ_n *and* Λ^f_n *are convex, compact and closed under the* w*-weak topology for each* $n \in \overline{\mathbb{N}}$.

Proof Under Assumptions 2.1, 3.1 and 3.3, it follows from Lemma 8.12 in [16] that Λ_n is convex and compact. Let $\{\mu_m\} \subset \Lambda_n$ with $\mu_m \xrightarrow{w} \mu$ and $v \in \mathbb{B}_b(S_n)$, by the definition of w-weakly convergence and Assumption 3.1(b), it follows from Lemma 8.11 in [16] that

$$\lim_{m\to\infty}\int_{K_n}\sum_{j\in S_n}v(j)q_n(j|i,a)\mu_m(i,\mathrm{d}a)=\int_{K_n}\sum_{j\in S_n}v(j)q_n(j|i,a)\mu(i,\mathrm{d}a)=0,$$

which implies that $\mu\in\Lambda_n$. Hence, Λ_n is closed. Furthermore, suppose that $\mu_m\in\Lambda_n^f$ for each m, then it follows that

$$\lim_{m\to\infty}\int_{K_n}c_n^l(i,a)\mu_m(i,\mathrm{d}a)=\int_{K_n}c_n^l(i,a)\mu(i,\mathrm{d}a)\le d_n^l,$$

for each $1\le l\le p$, which implies that $\mu\in\Lambda_n^f$ and Λ_n^f is closed. Now, suppose that μ_1, $\mu_2\in\Lambda_n^f$, we have that

$$\begin{aligned}&\int_{K_n}c_n^l(i,a)[\lambda\mu_1(i,\mathrm{d}a)+(1-\lambda)\mu_2(i,\mathrm{d}a)]\\&=\lambda\int_{K_n}c_n^l(i,a)\mu_1(i,\mathrm{d}a)+(1-\lambda)\int_{K_n}c_n^l(i,a)\mu_2(i,\mathrm{d}a)\le d_n^l,\end{aligned}$$

for each $1\le l\le p$, which implies the convexity of Λ_n^f. The proof is completed.

Lemma 3.5[10] *Suppose that Assumptions* 2.1, 3.1 *and* 3.3 *hold. Then, for each* $\pi\in\Pi_\infty$, *there exists a stationary policy* $\widetilde{\varphi}\in\Pi_\infty^s$ *such that* $J_\infty^0(\gamma_\infty,\widetilde{\varphi})\ge J_\infty^0(\gamma_\infty,\pi)$ *and* $J_\infty^l(\gamma_\infty,\widetilde{\varphi})\le J_\infty^l(\gamma_\infty,\pi)$ *for each* $1\le l\le p$.

By Assumption 3.4 and Lemma 3.5, there exists a stationary policy $\widetilde{\varphi}\in\Pi_\infty^s$ such that $J_\infty^l(\gamma_\infty,\widetilde{\varphi})<d_\infty^l$ for each $1\le l\le p$. Next, we need to introduce some notations:

$$g_n(\varphi):=(g_n^1(\varphi),\cdots,g_n^p(\varphi)),\quad \theta:=\min_{1\le l\le p}\{d_\infty^l-g_\infty^l(\widetilde{\varphi})\},\tag{3.2}$$

$$\epsilon_2(n):=\max_{1\le l\le p}\{|d_\infty^l-d_n^l|\},\quad \epsilon_1(n):=\max_{0\le l\le p}\{\sup_{\varphi\in\Pi_\infty^s}|g_\infty^l(\varphi)-g_n^l(\varphi|_{S_n})|\}\tag{3.3}$$

for each $n\in\overline{\mathbb{N}}$ and $\varphi\in\Pi_n^s$.

Lemma 3.6 *Suppose that Assumptions* 2.1 *and* 3.1-3.4 *hold, then there exist an integer* N^* *and* $\varphi_n\in\Pi_n^s$ *for each* $n\ge N^*$ *such that*

$$J_n^l(\gamma_n,\varphi_n)<d_n^l,\quad \textit{for any } 1\le l\le p.\tag{3.4}$$

Proof For each fixed $\delta<\theta$, it follows from Lemma 3.2 and Assumption 3.2(c) that there exists an integer N such that for each $n\ge N$, we have $\epsilon_1(n)+\epsilon_2(n)\le\delta$ and

$$g_\infty^l(\widetilde{\varphi})-\epsilon_1(n)\le g_n^l(\widetilde{\varphi}|_{S_n})\le g_\infty^l(\widetilde{\varphi})+\epsilon_1(n)<d_\infty^l-\epsilon_2(n)\le d_n^l,$$

for each $1\le l\le p$. Hence,

$$\begin{aligned}d_n^l-g_n^l(\widetilde{\varphi}|_{S_n})&=d_n^l-d_\infty^l+d_\infty^l-g_\infty^l(\widetilde{\varphi})+g_\infty^l(\widetilde{\varphi})-g_n^l(\widetilde{\varphi}|_{S_n})\\&\ge\theta-\epsilon_1(n)-\epsilon_2(n)>\theta-\delta>0,\end{aligned}\tag{3.5}$$

that is $g_n^l(\widetilde{\varphi}|_{S_n})<d_n^l$ for each $1\le l\le p$ which implies that $\varphi_n:=\widetilde{\varphi}|_{S_n}$ is a feasible solution of $\mathcal{M}_n$ for each $n\ge N^*$. The proof is completed.

The following lemma gives the optimality equation for constrained problem which has been established by Theorem 8.13 in [16] for the case of a single constraint. For completeness, we extend Theorem 8.13 in [16] to any finite numerable of constraints. First, we need to introduce some notations. For each $x = \{x_1, \cdots, x_p\}$, $y = \{y_1, \cdots, y_p\} \in \mathbb{R}^p$, we define $\langle x, y\rangle := \sum_{k=1}^{p} x_k \cdot y_k$ and say $x \le y$ if $x_k \le y_k$ for each $1 \le k \le p$.

The proof of the following lemma is similar to that of Theorem 4.10 in [5] for discrete-time MDP, we prove it here only for completeness.

Lemma 3.7 *Suppose that Assumptions* 2.1 *and* 3.1-3.4 *hold. Then, for each* $n \in \overline{\mathbb{N}}$,

(i) *there exist a function* $h_n \in B_w(S_n)$ *and a vector* $\lambda_n^* \in (-\infty, 0]^p$ *such that*

$$\begin{aligned} J_n^* &= \sup_{a\in A(i)} \Big\{ r_n(i,a) + \langle \lambda_n^*, c_n(i,a) - d_n\rangle + \sum_{j\in S_n} h_n(j) q_n(j|i,a) \Big\} \\ &= \sup_{\varphi\in\Pi_n^s} \Big\{ \int_{K_n} [r_n(i,a) + \langle \lambda_n^*, c_n(i,a) - d_n\rangle] \widetilde{\mu}_n^{\varphi}(i, \mathrm{d}a) \Big\} \\ &= \sup_{\mu\in\Lambda_n^f} \Big\{ \int_{K_n} r_n(i,a)\mu(i,\mathrm{d}a) \Big\}, \end{aligned}$$

where $c_n(i,a) - d_n$ *denotes the* p*-dimensional vector whose components are* $c_n^l(i,a) - d_n^l$ *for each* $1 \le l \le p$*;*

(ii) *there exists a stationary optimal policy* φ_n^* *of* $\mathcal{M}_n$.

Proof (i) Let $n \in \overline{\mathbb{N}}$ be fixed and

$$O_n := \bigcup_{\mu\in\Lambda_n} \Big\{ (z^1, \cdots, z^p) : \int_{K_n} c_n^l(i,a)\mu(i,\mathrm{d}a) \le z^l, \text{ for each } 1 \le l \le p \Big\}.$$

For each $z = (z^1, z^2, \cdots, z^p) \in O_n$ we define

$$\mathcal{N}_n(z) := \Big\{ \mu \in \Lambda_n : \int_{K_n} c_n^l(i,a)\mu(i,\mathrm{d}a) \le z^l, 1 \le l \le p \Big\}$$

and

$$U_n(z) := \sup \Big\{ \int_{K_n} r_n(i,a)\mu(i,\mathrm{d}a) : \mu \in \mathcal{N}_n(z) \Big\}.$$

Suppose that $\mu_m \xrightarrow{w} \mu$ and $\{\mu_m\} \subset \mathcal{N}_n(z)$, by the definition of w-weakly convergence and Assumption 3.1(b), then it follows that $\lim_{m\to\infty} \int_{K_n} c_n^l(i,a)\mu_m(i,\mathrm{d}a) = \int_{K_n} c_n^l(i,a)\mu(i,\mathrm{d}a) \le z^l$ for each $1 \le l \le p$ which implies that $\mu \in \mathcal{N}_n(z)$. Hence, $\mathcal{N}_n(z)$ is a closed set of Λ_n for each $n \in \overline{\mathbb{N}}$ and $z \in O_n$, which together with Lemma 3.4 implies $\mathcal{N}_n(z)$ is compact. Therefore, for each $z \in O_n$, there exists $\mu^* \in \mathcal{N}_n(z)$ such that $U_n(z) = \int_{K_n} r_n(i,a)\mu^*(i,\mathrm{d}a)$ which together with Lemma 3.1 implies $U_n(z)$ is finite for each $z \in O_n$.

Next, we need to show that $U_n(z)$ is concave in z. Let $z_k \in O_n$ with $z_k = (z_k^1, \cdots, z_k^p)$ for each $k = 1, 2$. There exist $\mu_1 \in \mathcal{N}_n(z_1)$ and $\mu_2 \in \mathcal{N}_n(z_2)$ such that $\int_{K_n} r_n(i,a)\mu_1(i,\mathrm{d}a) = U_n(z_1)$ and $\int_{K_n} r_n(i,a)\mu_2(i,\mathrm{d}a) = U_n(z_2)$. Let $\lambda \in (0,1)$. We have

$$\lambda U_n(z_1) + (1-\lambda)U_n(z_2) = \int_{K_n} r_n(i,a)[\lambda\mu_1 + (1-\lambda)\mu_2](i,\mathrm{d}a)$$

and

$$\int_{K_n} c_n^l(i,a)[\lambda\mu_1 + (1-\lambda)\mu_2](i,\mathrm{d}a) \le \lambda z_1^l + (1-\lambda)z_2^l, \quad \text{for each } 1 \le l \le p,$$

which implies that $\lambda\mu_1 + (1-\lambda)\mu_2 \in \mathcal{N}_n(\lambda z_1 + (1-\lambda)z_2)$. Thus, it follows that $U_n(\lambda z_1 + (1-\lambda)z_2) \ge \lambda U_n(z_1) + (1-\lambda)U_n(z_2)$, that is $U_n(z)$ is concave on O_n.

For arbitrary $z_1, z_2 \in O_n$ satisfying $z_1^l \le z_2^l$ for each $1 \le l \le p$, we have $U_n(z_1) \le U_n(z_2)$. By Lemma 3.6, d_n is the interior of O_n. Then, it follows from Theorem 7.12 in [1] that there exists $\lambda_n^* = (\lambda_n^{*1}, \cdots, \lambda_n^{*p}) \in \mathbb{R}^p$ with $\lambda_n^* \le 0$ such that $U_n(\hat{\theta}_n) \le U_n(d_n) + \langle -\lambda_n^*, \hat{\theta}_n - d_n\rangle$, for each $\hat{\theta}_n \in O_n$. Take $\hat{\theta}_n := \{\int_{K_n} c_n^1(i,a)\mu(i,\mathrm{d}a), \cdots, \int_{K_n} c_n^p(i,a)\mu(i,\mathrm{d}a)\}$ for some $\mu \in \Lambda_n$. Then,

$$\begin{aligned} U_n(d_n) \ge U_n(\hat{\theta}_n) + \langle \lambda_n^*, \hat{\theta}_n - d_n\rangle &= \int_{K_n} r_n(i,a)\mu(i,\mathrm{d}a) + \langle \lambda_n^*, \hat{\theta}_n - d_n\rangle \\ &= \int_{K_n} r_n(i,a)\mu(i,\mathrm{d}a) + \int_{K_n} \langle \lambda_n^*, c_n(i,a) - d_n\rangle \mu(i,\mathrm{d}a), \end{aligned}$$

for each $\mu \in \Lambda_n$. That is,

$$\begin{aligned} U_n(d_n) &\ge \sup_{\mu\in\Lambda_n} \Big\{ \int_{K_n} r_n(i,a)\mu(i,\mathrm{d}a) + \int_{K_n} \langle \lambda_n^*, c_n(i,a) - d_n\rangle \mu(i,\mathrm{d}a) \Big\} \\ &\ge \sup_{\mu\in\Lambda_n^f} \Big\{ \int_{K_n} r_n(i,a)\mu(i,\mathrm{d}a) \Big\} = U_n(d_n), \end{aligned}$$

which implies that

$$\begin{aligned} U_n(d_n) &= \sup_{\mu\in\Lambda_n} \Big\{ \int_{K_n} r_n(i,a)\mu(i,\mathrm{d}a) + \int_{K_n} \langle \lambda_n^*, c_n(i,a) - d_n\rangle \mu(i,\mathrm{d}a) \Big\} \\ &= \sup_{\varphi\in\Pi_n^s} \Big\{ \int_{K_n} \big(r_n(i,a) + \langle \lambda_n^*, c_n(i,a) - d_n\rangle\big) \widetilde{\mu}_n^{\varphi}(i,\mathrm{d}a) \Big\}. \end{aligned} \tag{3.6}$$

Under Assumptions 2.1, 3.1 and 3.3, by (3.6), Theorem 3.20 in [16] and Remark 3.2(ii), there exists $h_n \in B_w(S_n)$ such that

$$\begin{aligned} U_n(d_n) &= \max_{a\in A(i)} \Big\{ r_n(i,a) + \langle \lambda_n^*, c_n(i,a) - d_n\rangle + \sum_{j\in S_n} h_n(j) q_n(j|i,a) \Big\} \\ &= \sup_{\pi\in\Pi_n} \Big\{ \liminf_{T\to\infty} \frac{1}{T} E_{\gamma_n,n}^{\pi} \Big[\int_0^T \int_A \Big(r_n(\xi_t,a) + \sum_{l=1}^{p} \lambda_n^{*l} (c_n^l(\xi_t,a) - d_n^l) \Big) \pi(\mathrm{d}a|\xi_t,t)\mathrm{d}t \Big] \Big\}. \end{aligned}$$

Let $\pi \in \Pi_n$ such that

$$\limsup_{T\to\infty} \frac{1}{T} E^{\pi}_{\gamma_n,n}\Big[\int_0^T\int_A c_n^l(\xi_t,a)\pi(\mathrm{d}a|\xi_t,t)\mathrm{d}t\Big] \le d_n^l,$$

for each $1 \le l \le p$, then we have

$$\begin{aligned} J_n^0(\gamma_n,\pi) &= \liminf_{T\to\infty} \frac{1}{T} E^{\pi}_{\gamma_n,n}\Big[\int_0^T\int_A r_n(\xi_t,a)\pi(\mathrm{d}a|\xi_t,t)\mathrm{d}t\Big] \\ &\le \liminf_{T\to\infty} \frac{1}{T} E^{\pi}_{\gamma_n,n}\Big[\int_0^T\int_A \Big(r_n(\xi_t,a) + \sum_{l=1}^{p} \lambda_n^{*l}(c_n^l(\xi_t,a) - d_n^l)\Big)\pi(\mathrm{d}a|\xi_t,t)\mathrm{d}t\Big] \\ &\le U_n(d_n). \end{aligned}$$

On the other hand, by the definition of $U_n(d_n)$, we have $J_n^* \ge U_n(d_n)$. Hence, $J_n^* = U_n(d_n)$. (ii) Let $n \in \overline{\mathbb{N}}$. By Assumption 3.1(b) and Lemma 3.4, there exist a p.m. $\mu^* \in \Lambda_n^f$ and a stationary policy $\varphi_n^* \in \Pi_n^s$ such that

$$J_n^* = \int_{K_n} r_n(i,a)\mu^*(i,\mathrm{d}a) \quad \text{and} \quad \mu^*(i,\mathrm{d}a) = \widetilde{\mu}_n^{\varphi_n^*}(i,\mathrm{d}a),$$

which implies that φ_n^* is optimal for $\mathcal{M}_n$. The proof is completed.

The proof of the following lemma is similar to that of Theorem 8.14 in [16] with one constraint, we prove it here for the sake of completeness.

Lemma 3.8 *Suppose that Assumptions* 2.1 *and* 3.1-3.4 *hold. Then, for each* $n \in \overline{\mathbb{N}}$*, there exist* $\mu_n^* \in \Lambda_n$ *and* $\lambda_n^* \in (-\infty, 0]^p$ *such that*

$$\begin{aligned} J_n^* &= \sup_{\mu\in\Lambda_n} \inf_{\lambda\in(-\infty,0]^p} \Big\{\int_{K_n} [r_n(i,a) + \langle\lambda, c_n(i,a) - d_n\rangle]\mu(i,\mathrm{d}a)\Big\} \\ &= \inf_{\lambda\in(-\infty,0]^p} \sup_{\mu\in\Lambda_n} \Big\{\int_{K_n} [r_n(i,a) + \langle\lambda, c_n(i,a) - d_n\rangle]\mu(i,\mathrm{d}a)\Big\} \\ &= \sup_{\mu\in\Lambda_n} \Big\{\int_{K_n} [r_n(i,a) + \langle\lambda_n^*, c_n(i,a) - d_n\rangle]\mu(i,\mathrm{d}a)\Big\} \\ &= \inf_{\lambda\in(-\infty,0]^p} \Big\{\int_{K_n} [r_n(i,a) + \langle\lambda, c_n(i,a) - d_n\rangle]\mu_n^*(i,\mathrm{d}a)\Big\}. \end{aligned}$$

Proof Let λ_n^* be the vector as in Lemma 3.7, we can get the third equality. For each fixed $n \in \overline{\mathbb{N}}$, suppose that $\mu \in \Lambda_n$ such that the l-th constraints is violated, that is $\int_{K_n} c_n^l(i,a)\mu(i,\mathrm{d}a) > d_n^l$. It follows that

$$\inf_{\lambda\in(-\infty,0]^p} \Big\{\int_{K_n} [r_n(i,a) + \langle\lambda, c_n(i,a) - d_n\rangle]\mu(i,\mathrm{d}a)\Big\} = -\infty.$$

Hence,

$$\begin{aligned}
&\sup_{\mu\in\Lambda_n}\inf_{\lambda\in(-\infty,0]^p}\Big\{\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu(i,\mathrm{d}a)\Big\}\\
=&\sup_{\mu\in\Lambda_n^f}\inf_{\lambda\in(-\infty,0]^p}\Big\{\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu(i,\mathrm{d}a)\Big\}\\
=&\sup_{\mu\in\Lambda_n^f}\Big\{\int_{K_n}r_n(i,a)\mu(i,\mathrm{d}a)\Big\},
\end{aligned}$$

which together with Lemma 3.7 implies the first equality.

Let $T_n(\mu,\lambda):=\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu(i,\mathrm{d}a)$ for each $\mu\in\Lambda_n$ and $\lambda\in(-\infty,0]^p$. For arbitrary fixed $\lambda\in(-\infty,0]^p$, $\beta\in(0,1)$ and $\mu_1,\mu_2\in\Lambda_n$, we have that

$$\begin{aligned}
&\beta T_n(\mu_1,\lambda)+(1-\beta)T_n(\mu_2,\lambda)\\
=&\beta\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu_1(i,\mathrm{d}a)\\
&+(1-\beta)\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu_2(i,\mathrm{d}a)\\
=&\int_{K_n}\big(r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle\big)[\beta\mu_1+(1-\beta)\mu_2](i,\mathrm{d}a)\\
=&T_n(\beta\mu_1+(1-\beta)\mu_2,\lambda).
\end{aligned}$$

Hence, $T_n(\mu,\lambda)$ is convex in μ for each fixed λ.

Let $f(i,a):=r_n(i,a)+\sum_{l=1}^{p}\lambda^l(c_n^l(i,a)-d_n^l)$, then

$$|f(i,a)|\le Mw(i)+p|\lambda|_{\max}[Mw(i)+d_n^{\max}]\le M\Big(1+p|\lambda|_{\max}+\frac{p|\lambda|_{\max}d_n^{\max}}{M}\Big)w(i),$$

where $|\lambda|_{\max}:=\max_{1\le l\le p}\{|\lambda^l|\}$ and $d_n^{\max}:=\max_{1\le l\le p}\{|d_n^l|\}$, which together with Assumption 3.1(b) implies that $f\in\mathcal{C}_w(K_n)$. Now, suppose that $\{\mu_m\}\subseteq\Lambda_n$ such that $\mu_m\xrightarrow{w}\mu$, we have $\mu\in\Lambda_n$ and

$$\lim_{m\to\infty}\int_{K_n}f(i,a)\mu_m(i,\mathrm{d}a)=\int_{K_n}f(i,a)\mu(i,\mathrm{d}a),$$

which implies that $T_n(\mu,\lambda)$ is upper semi-continuous in μ for each fixed λ.

For each fixed $\mu\in\Lambda_n$, let $\lambda_1,\lambda_2\in(-\infty,0]^p$. It follows that

$$\beta T_n(\mu,\lambda_1)+(1-\beta)T_n(\mu,\lambda_2)=\int_{K_n}[r_n(i,a)+\langle\beta\lambda_1+(1-\beta)\lambda_2,c_n(i,a)-d_n\rangle]\mu(i,\mathrm{d}a)$$

which implies that $T_n(\mu,\lambda)$ is concave in λ for each μ. By the Minmax Theorem in [2, p.129] or [7, Theorem 2], we have

$$\sup_{\mu\in\Lambda_n}\inf_{\lambda\in(-\infty,0]^p} T_n(\mu,\lambda)=\inf_{\lambda\in(-\infty,0]^p}\sup_{\mu\in\Lambda_n} T_n(\mu,\lambda),$$

that is

$$\begin{aligned}&\sup_{\mu\in\Lambda_n}\inf_{\lambda\in(-\infty,0]^p}\Big\{\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu(i,\mathrm{d}a)\Big\}\\ =&\inf_{\lambda\in(-\infty,0]^p}\sup_{\mu\in\Lambda_n}\Big\{\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu(i,\mathrm{d}a)\Big\},\end{aligned}$$

which implies the second equality. By Lemma 3.7, there exists a stationary policy $\varphi_n^*\in\Phi_n^s$ such that $J_n^*=J_n^0(\gamma_n,\varphi_n^*)$. Let $\mu_n^*:=\widetilde{\mu}_n^{\varphi_n^*}$, then

$$J_n^*=\int_{K_n} r_n(i,a)\mu_n^*(i,\mathrm{d}a)=\inf_{\lambda\in(-\infty,0]^p}\Big\{\int_{K_n}[r_n(i,a)+\langle\lambda,c_n(i,a)-d_n\rangle]\mu_n^*(i,\mathrm{d}a)\Big\}.$$

Hence, the last inequality holds. The proof of Lemma 3.8 is finished.

4 The Main Results

Theorem 4.1 *Suppose that Assumptions* 2.1 *and* 3.1-3.4 *hold. Then,* $\lim_{n\to\infty} J_n^*=J_\infty^*$.

Proof Let φ_n be the policy in Lemma 3.6, $B:=M\frac{\kappa_1'}{\rho_1'}$ and $\delta<\theta$ be fixed. By Lemmas 3.1 and 3.7 and (3.5),

$$\begin{aligned}B\geq J_n^*&=\sup_{\varphi\in\Pi_n^s}\Big\{\int_{K_n}[r_n(i,a)+\langle\lambda_n^*,c_n(i,a)-d_n\rangle]\widetilde{\mu}_n^{\varphi}(i,\mathrm{d}a)\Big\}\\ &\geq\int_{K_n}[r_n(i,a)+\langle\lambda_n^*,c_n(i,a)-d_n\rangle]\widetilde{\mu}_n^{\varphi_n}(i,\mathrm{d}a)\\ &\geq -B+\langle\lambda_n^*,g_n(\widetilde{\mu}_n^{\varphi_n})-d_n\rangle=-B+\langle-\lambda_n^*,d_n-g_n(\widetilde{\mu}_n^{\varphi_n})\rangle\\ &>-B+\langle-\lambda_n^*,\theta-\delta\rangle.\end{aligned}$$

Then, we have

$$\langle-\lambda_n^*,e\rangle\leq\frac{2B}{\theta-\delta},\quad\text{for any } n\geq N.\tag{4.1}$$

As in Lemma 3.6, there exists a stationary policy $\widetilde{\varphi}\in\Pi_\infty^s$ such that $J_\infty^l(\gamma_\infty,\widetilde{\varphi})<d_\infty^l$ for each $1\leq l\leq p$. Similarly, we have

$$\begin{aligned}B\geq J_\infty^*&=\sup_{\varphi\in\Pi_\infty^s}\Big\{\int_{K_\infty}[r_\infty(i,a)+\langle\lambda_\infty^*,c_\infty(i,a)-d_\infty\rangle]\widetilde{\mu}_\infty^{\varphi}(i,\mathrm{d}a)\Big\}\\ &\geq\int_{K_\infty}[r_\infty(i,a)+\langle\lambda_\infty^*,c_\infty(i,a)-d_\infty\rangle]\widetilde{\mu}_\infty^{\widetilde{\varphi}}(i,\mathrm{d}a)\\ &\geq -B+\langle-\lambda_\infty^*,d_\infty-g_\infty(\widetilde{\varphi})\rangle>-B+\langle-\lambda_\infty^*,\theta\rangle,\end{aligned}$$

which implies that

$$\langle -\lambda_\infty^*, e\rangle \le \frac{2B}{\theta} \le \frac{2B}{\theta-\delta}. \tag{4.2}$$

Under Assumptions 2.1, 3.1 and 3.3, by Lemma 3.8, there exists $\mu_\infty^* \in \Lambda_\infty$ such that

$$J_\infty^* \le \int_{K_\infty} [r_\infty(i,a) + \langle \lambda_n^*, c_\infty(i,a) - d_\infty\rangle]\mu_\infty^*(i,\mathrm{d}a). \tag{4.3}$$

By (4.1) and (4.3), there exists $\varphi \in \Pi_\infty^s$ such that $\mu_\infty^* = \widetilde{\mu}_\infty^\varphi$. Let $\hat{\varphi}_n := \varphi|_{S_n}$ denote the restriction of φ in the set S_n, then

$$\begin{aligned}
&J_\infty^* - J_n^* \\
\le &\int_{K_\infty} [r_\infty(i,a) + \langle \lambda_n^*, c_\infty(i,a) - d_\infty\rangle]\mu_\infty^*(i,\mathrm{d}a) \\
&- \sup_{\mu\in\Lambda_n} \Big\{ \int_{K_n} [r_n(i,a) + \langle \lambda_n^*, c_n(i,a) - d_n\rangle]\mu(i,\mathrm{d}a)\Big\} \\
\le &\int_{K_\infty} [r_\infty(i,a) + \langle \lambda_n^*, c_\infty(i,a) - d_\infty\rangle]\mu_\infty^*(i,\mathrm{d}a) \\
&- \int_{K_n} [r_n(i,a) + \langle \lambda_n^*, c_n(i,a) - d_n\rangle]\widetilde{\mu}_n^{\hat{\varphi}_n}(i,\mathrm{d}a) \\
= &\, g_\infty^0(\varphi) + \langle \lambda_n^*, g_\infty(\varphi) - d_\infty\rangle - g_n^0(\hat{\varphi}_n) - \langle \lambda_n^*, g_n(\hat{\varphi}_n) - d_n\rangle \\
\le &\, \epsilon_1(n) + \langle \lambda_n^*, g_\infty(\varphi) - g_n(\hat{\varphi}_n)\rangle + \langle \lambda_n^*, d_n - d_\infty\rangle \\
\le &\, \epsilon_1(n) + \langle -\lambda_n^*, \epsilon_1(n)\rangle + \langle -\lambda_n^*, \epsilon_2(n)\rangle = \epsilon_1(n) + [\epsilon_1(n) + \epsilon_2(n)]\frac{2B}{\theta-\delta}.
\end{aligned}$$

Similarly, by Lemma 3.8, there exists $\mu_n^* \in \Lambda_n$ such that

$$J_n^* \le \int_{K_n} [r_n(i,a) + \langle \lambda_\infty^*, c_n(i,a) - d_n\rangle]\mu_n^*(i,\mathrm{d}a). \tag{4.4}$$

Then, there exists $\varphi_n \in \Pi_n^s$ such that $\mu_n^* = \widetilde{\mu}_n^{\varphi_n}$ for each $n \in \mathbb{N}$. Let $n \in \mathbb{N}$ be fixed and $\widetilde{\varphi}_n$ denote an extension of φ_n to Π_∞^s by:

$$\widetilde{\varphi}_n(\cdot|i) := \begin{cases} \varphi_n(\cdot|i) & \text{if } i \in S_n, \\ \nu & \text{if } i \in S_n^c, \text{ where } \nu \in \mathcal{P}(A(i)) \text{ is chosen arbitrarily.} \end{cases} \tag{4.5}$$

It follows from (4.2) and (4.4)-(4.5) that

$$\begin{aligned}
&J_n^* - J_\infty^* \\
= &\inf_{\lambda\in(-\infty,0]^p} \int_{K_n} [r_n(i,a) + \langle \lambda, c_n(i,a) - d_n\rangle]\mu_n^*(i,\mathrm{d}a) \\
&- \sup_{\mu\in\Lambda_\infty} \int_{K_\infty} [r_\infty(i,a) + \langle \lambda_\infty^*, c_\infty(i,a) - d_\infty\rangle]\mu(i,\mathrm{d}a)
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_{K_n} [r_n(i,a) + \langle \lambda^*_\infty, c_n(i,a) - d_n\rangle]\mu^*_n(i,\mathrm{d}a) \\
&\quad - \int_{K_\infty} [r_\infty(i,a) + \langle \lambda^*_\infty, c_\infty(i,a) - d_\infty\rangle]\widetilde{\mu}^{\widetilde{\varphi}_n}_\infty(i,\mathrm{d}a) \\
&= g^0_n(\varphi_n) - g^0_\infty(\widetilde{\varphi}_n) + \langle \lambda^*_\infty, g_n(\varphi_n) - g_\infty(\widetilde{\varphi}_n) + d_\infty - d_n\rangle \\
&\leq \epsilon_1(n) + \langle -\lambda^*_\infty, g_\infty(\widetilde{\varphi}_n) - g_n(\varphi_n)\rangle + \langle -\lambda^*_\infty, d_n - d_\infty\rangle \\
&\leq \epsilon_1(n) + [\epsilon_1(n) + \epsilon_2(n)]\langle -\lambda^*_\infty, e\rangle \\
&\leq \epsilon_1(n) + [\epsilon_1(n) + \epsilon_2(n)]\frac{2B}{\theta - \delta}.
\end{aligned}
$$

Hence, there exists an integer N such that for each $n \geq N$, $|J^*_n - J^*_\infty| \leq \epsilon_1(n) + [\epsilon_1(n) + \epsilon_2(n)]\frac{2B}{\theta-\delta}$, which implies the desired result. The proof is completed.

Lemma 4.1[11] *Suppose that Assumptions* 2.1, 3.1 *and* 3.3 *hold. For any fixed* $n \in \overline{\mathbb{N}}$, *if there exists* $\{\varphi_m\} \subseteq \Pi^s_n$ *such that* $\varphi_m \to \varphi \in \Pi^s_n$, *then* $\lim\limits_{m\to\infty} g^l_n(\varphi_m) = g^l_n(\varphi)$ *for each* $0 \leq l \leq p$.

Theorem 4.2 *Suppose that Assumptions* 2.1 *and* 3.1-3.4 *hold. If* $\varphi^*_n \in \Pi^s_n$ *is an optimal policy of* $\mathcal{M}_n$ *for each* $n \in \mathbb{N}$ *and* $\varphi^*_n \to \varphi_\infty \in \Pi^s_\infty$, *then* φ_∞ *is an optimal policy of* $\mathcal{M}_\infty$.

Proof First, we should show that φ_∞ is a feasible solution of $\mathcal{M}_\infty$. Let $\widetilde{\varphi}^*_n$ denote an extension of φ^*_n to Π^s_∞ by replacing φ_n in (4.5) with φ^*_n here. Under Assumptions 2.1 and 3.1-3.3, by Lemmas 3.2 and 4.1, for each $\epsilon > 0$, there exists an integer N such that $|g^l_n(\varphi^*_n) - g^l_\infty(\widetilde{\varphi}^*_n)| \leq \frac{\epsilon}{2}$ and $|g^l_\infty(\widetilde{\varphi}^*_n) - g^l_\infty(\varphi_\infty)| \leq \frac{\epsilon}{2}$ for each $n \geq N$ and $1 \leq l \leq p$. Hence,

$$
|g^l_n(\varphi^*_n) - g^l_\infty(\varphi_\infty)| = |g^l_n(\varphi^*_n) - g^l_\infty(\widetilde{\varphi}^*_n) + g^l_\infty(\widetilde{\varphi}^*_n) - g^l_\infty(\varphi_\infty)| \leq \epsilon,
$$

which together with Assumption 3.2(c) implies that

$$
d^l_\infty = \lim_{n\to\infty} d^l_n \geq \lim_{n\to\infty} \int_{K_n} c^l_n(i,a)\widetilde{\mu}^{\varphi^*_n}_n(i,\mathrm{d}a) = \int_{K_\infty} c^l_\infty(i,a)\widetilde{\mu}^{\varphi_\infty}_\infty(i,\mathrm{d}a),
$$

for each $1 \leq l \leq p$. Hence, φ_∞ is feasible for $\mathcal{M}_\infty$. By Lemma 4.1 and Theorem 3.6, we have that

$$
\begin{aligned}
g^0_\infty(\varphi_\infty) &= \lim_{n\to\infty} g^0_\infty(\widetilde{\varphi}^*_n) \\
&\geq \overline{\lim_{n\to\infty}}[g^0_n(\varphi^*_n) - \sup_{\varphi\in\Pi^s_\infty} |g^0_n(\varphi|_{S_n}) - g^0_\infty(\varphi)|] \\
&= \overline{\lim_{n\to\infty}} J^*_n - \underline{\lim_{n\to\infty}} \sup_{\varphi\in\Pi^s_\infty} |g^0_n(\varphi|_{S_n}) - g^0_\infty(\varphi)| = J^*_\infty.
\end{aligned}
$$

Hence, φ_∞ is optimal for $\mathcal{M}_\infty$. The proof is completed.

References

[1] C.D. Aliprantis, and K.C. Bordr, Infinite Dimensional Analysis, Springer, New York, 2006.

[2] E. Altman, Constrained Markov Decision Processes, Chapman Hall/CRC, Boca Raton, FL, 1999.

[3] J. Alvarez-Mena, and O. Hernández-Lerma, Convergence of the optimal values of constrained Markov control processes, *Math. Methods Oper. Res.*, **55**(2002),461-484.

[4] V.S. Borkar, An actor-critic algorithm for constrained Markov decision processes, *Systems Control Lett.*, **54**(2005),207-213.

[5] F. Dufour, and T. Prieto-Rumeau, Finite linear programming approximations of constrained discounted Markov decision processes, *SIAM J. Control Optim.*, **51**(2013), 1298-1324.

[6] F. Dufour, and T. Prieto-Rumeau, Stochastic approximations of constrained discounted Markov decision processes, *J. Math. Anal. Appl.*, **413**(2014),856-879.

[7] K. Fan, Minimax theorems, *Proc. Nat. Acad. Sci. U.S.A.*, **39**(1953),42-47.

[8] H. Föllmer, and A. Schied, Stochastic Finance: An Introduction in Discrete Time, Walter de Gruyter, Berlin, 2004.

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

[10] X.P. Guo, Q.D. Wei and J.Y. Zhang, A constrained optimization problem with applications to constrained MDPs, *Optimization, Control, and Applications of Stochastic Systems*, 125-150, Systems Control Found. Appl., Birkhäuser/Springer, New York, 2012.

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

[12] X.P Guo, and W.Z. Zhang, Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints, *European J. Oper. Res.*, **238**(2014),486-496.

[13] O. Hernández-Lerma, and J.B. Lasserre, Discrete-Time Markov Control Processes: Basic Optimality Criteria, Springer-Verlag, New York, 1996.

[14] M.Y. Kitaev, and V.V. Rykov, Controlled Queueing Systems, CRC Press, Boca Raton, 1995.

[15] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley, New York, 1994.

[16] T. Prieto-Rumeau, and O. Hernández-Lerma, Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games, World Scientific, 2012.

[17] Q.D. Wei, Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion, *Math. Methods Oper. Res.*, **84**(2016),461-487.

[18] Q.D. Wei, Finite approximation for finite-horizon continuous-time Markov decision processes, *4OR*, **15**(2017),67-84.

(*edited by Liangwei Huang*)