Chapter 8
A Constrained Optimization Problem with Applications to Constrained MDPs

Xianping Guo, Qingda Wei, and Junyu Zhang

8.1 Introduction

Constrained optimization problems form an important aspect in control theory, for instance, constrained Markov decision processes (MDPs) [2, 3, 10–13, 15–21, 24, 25, 27, 29–33, 35, 36], and constrained diffusion processes [4–9]. In this chapter, we are concerned with a constrained optimization problem, in which the objective function is defined on the product space of a linear space and a convex set. The constrained optimization problem is to maximize the values of the function with any fixed variable in the linear space, over a constrained subset of the convex set which is given by the function with *another* fixed variable from the linear space and with a given constraint. The basic idea for the constrained optimization problem comes from the studies on the discounted and average optimality for discrete- and continuous-time MDPs with a constraint. We aim to develop a *unified* approach to dealing with such constrained MDPs. More precisely, for discrete- and continuous-time MDPs with a constraint, the linear space can be taken as a set of some real-valued functions such as reward and cost functions in such MDPs, and the convex set can be chosen as the set of all randomized Markov policies, the set of all randomized stationary policies, or the set of all the occupation measures according to a specified case of MDPs with different criteria. The objective function can be taken as one of the expected discounted (average) criteria, while the first variable in the objective function can be taken as the reward/cost functions in MDPs and the second one as a policy in a class of policies. Thus, MDPs with a constraint can be reduced to our constrained optimization problem.

Research supported by NSFC, GDUPS, RFDP, and FRFCU.

X. Guo (✉) • Q. Wei • J. Zhang
Sun Yat-Sen University, Guangzhou 510275, China
e-mail: mcsgxp@mail.sysu.edu.cn; wwqingda@sina.com; mcszhjy@mail.sysu.edu.cn

D. Hernández-Hernández and A. Minjárez-Sosa (eds.), *Optimization, Control, and Applications of Stochastic Systems*, Systems & Control: Foundations & Applications,
DOI 10.1007/978-0-8176-8337-5_8,

A fundamental question on the constrained optimization problem is whether there exists a constrained-optimal solution. The Lagrange multiplier technique is a classical approach to proving the existence of a constrained-optimal solution for such an optimization problem. There are several authors using the Lagrange multiplier technique to study MDPs with a constraint; see, for instance, discrete-time constrained MDPs with the discounted and average criteria [3, 32, 33] and continuous-time constrained MDPs with the discounted and average criteria [18, 20, 35]. All the aforementioned works [3, 18, 20, 32, 33, 35] require the nonnegativity assumption on the costs. We also apply this approach to analyze the constrained optimization problem. Following the arguments in [3, 18, 20, 32, 33, 35], we give conditions under which we prove the existence of a constrained-optimal solution to the constrained optimization problem, and also give a characterization of a constrained-optimal solution for a particular case.

Then, we apply our main results to discrete- and continuous-time constrained MDPs with discounted and average criteria. More precisely, in Sect. 8.4.1, we use the results to show the existence of a constrained-optimal policy for the discounted discrete-time MDPs with a constraint in which the state space is a Polish space and the rewards/costs may be unbounded from above and from below. To the best of our knowledge, there are no any existing works dealing with constrained discounted discrete-time MDPs in Borel spaces and with unbounded rewards/costs. In Sect. 8.4.2, we investigate an application of the main results to constrained discrete-time MDPs with state-dependent discount factors and extend the results in [32] to the case in which discount factors can depend on states and rewards/costs can be unbounded from above and from below. In Sects. 8.4.3 and 8.4.4, we consider the average and discounted continuous-time MDPs with a constraint, respectively. Removing the nonnegativity assumption on the cost function as in [18, 20, 35], we prove that the results in [18, 20, 35] still hold using the results in this chapter.

The rest of this chapter is organized as follows. In Sect. 8.2, we introduce the constrained optimization problem under consideration and give some preliminary facts needed to prove the existence of a constrained-optimal solution to the optimization problem. In Sect. 8.3, we state and prove our main results on the existence of a constrained-optimal solution. In Sect. 8.4, we provide some applications of our main results to constrained MDPs with different optimality criteria.

8.2 A Constrained Optimization Problem

In this section, we state the constrained optimization problem under consideration and give some preliminary results needed to prove the existence of a constrained-optimal solution. To do so, we introduce some notation below:

1. Let C be a linear space and D a convex set.

2. Suppose that G is a real-valued function on the product space $C \times D$ and satisfies the following property:

$$G(k_1c_1 + k_2c_2, d) = k_1G(c_1, d) + k_2G(c_2, d) \tag{8.1}$$

for any $c_1, c_2 \in C$, $d \in D$ and any constants $k_1, k_2 \in \mathbb{R} := (-\infty, +\infty)$.

For any *fixed* $c \in C$, let

$$U := \{d \in D : G(c,d) \leq \rho\},$$

which depends on the given c and a so-called constraint constant ρ.

Then, for another given $r \in C$, we consider a constrained optimization problem below:

$$\text{Maximize } G(r, \cdot) \text{ over } U. \tag{8.2}$$

Definition 8.2.1. $d^* \in U$ is said to be a constrained-optimal solution to the problem (8.2) if d^* maximizes $G(r,d)$ over $d \in U$, that is,

$$G(r, d^*) = \sup_{d \in U} G(r,d).$$

Remark 8.2.1. When D is a compact and convex metric space, and $G(r,d)$ and $G(c,d)$ are continuous in $d \in D$, it follows from the Weierstrass theorem [1, p.40] that there exists a constrained-optimal solution. In general, however, D may be unmetrizable in some cases, such as the set of all randomized Markov policies in continuous-time MDPs [20, p.10]; see continuous-time constrained MDPs with the discounted criteria in Sect. 8.4.4. In order to solve (8.2), we *assume that there exists a subset $D' \subseteq D$, which is assumed to be a compact metric space throughout this chapter*.

To analyze problem (8.2), we define the following unconstrained optimization problem by introducing a Lagrange multiplier $\lambda \geq 0$,

$$b^\lambda := r - \lambda c, \quad G^*(b^\lambda) := \sup_{d \in D} G(b^\lambda, d), \tag{8.3}$$

and then give the conditions below.

Assumption 8.2.1

(i) *For each fixed* $\lambda \geq 0$, $D^*_\lambda := \{d^\lambda \in D' \mid G(b^\lambda, d^\lambda) = G^*(b^\lambda)\} \neq \emptyset$.
(ii) *There exists a constant* $M > 0$, *such that* $\max\{|G(r,d)|, |G(c,d)|\} \leq M$ *for all* $d \in D$.
(iii) $G(r,d)$ *and* $G(c,d)$ *are continuous in* $d \in D'$.

Assumption 8.2.1(i) implies that there exists at least an element $d^\lambda \in D'$ such that $G(b^\lambda, \cdot)$ attains its maximum. In addition, the boundedness and continuity hypotheses in Assumptions 8.2.1(ii) and (iii) are commonly used in optimization control theory.

Assumption 8.2.2 *For the given $c \in C$, there exists an element $d' \in D$ (depending on c) such that $G(c,d') < \rho$, which means that $\{d \in D \mid G(c,d) < \rho\} \neq \emptyset$.*

Remark 8.2.2. Assumption 8.2.2 is a Slater-like hypothesis, typical for the constrained optimization problems; see, for instance, [3, 17, 18, 20, 32, 33, 35].

In order to prove the existence of a constrained-optimal solution, we need the following preliminary lemmas.

Lemma 8.2.1. *Suppose that Assumption 8.2.1(i) holds. Then, $G(c,d^\lambda)$ is nonincreasing in $\lambda \in [0,\infty)$, where $d^\lambda \in D^*_\lambda$ is arbitrary but fixed for each $\lambda \geq 0$.*

Proof. For each $d \in D$, by (8.1) and (8.3), we have

$$G(b^\lambda, d) = G(r,d) - \lambda G(c,d) \quad \text{for all } \lambda \geq 0.$$

Moreover, since $G(b^\lambda, d^\lambda) = G^*(b^\lambda)$ for all $\lambda \geq 0$ and $d^\lambda \in D^*_\lambda$, we have, for any $h > 0$,

$$\begin{aligned} -hG(c,d^\lambda) &= G(b^{\lambda+h}, d^\lambda) - G(b^\lambda, d^\lambda) \\ &\leq G(b^{\lambda+h}, d^{\lambda+h}) - G(b^\lambda, d^\lambda) \\ &\leq G(b^{\lambda+h}, d^{\lambda+h}) - G(b^\lambda, d^{\lambda+h}) \\ &= -hG(c,d^{\lambda+h}), \end{aligned}$$

which implies that

$$G(c,d^\lambda) \geq G(c,d^{\lambda+h}).$$

Hence, $G(c,d^\lambda)$ is nonincreasing in $\lambda \in [0,\infty)$. □

Remark 8.2.3. Under Assumption 8.2.1(i), it follows from Lemma 8.2.1 that the following nonnegative constant

$$\widetilde{\lambda} := \inf\{\lambda \geq 0 : G(c,d^\lambda) \leq \rho, d^\lambda \in D^*_\lambda\} \tag{8.4}$$

is well defined.

Lemma 8.2.2. *Suppose that Assumptions 8.2.1(i), (ii) and 8.2.2 hold. Then, the constant $\widetilde{\lambda}$ in (8.4) is finite; that is, $\widetilde{\lambda}$ is in $[0,\infty)$.*

Proof. Let $\kappa := \rho - G(c,d') > 0$, with d' as in Assumption 8.2.2. Since $\lim_{\lambda\to\infty} \frac{2M}{\lambda} = 0$ for the constant M as in Assumption 8.2.1(ii), there exists $\delta > 0$ such that

$$\frac{2M}{\lambda} - \kappa < 0 \text{ for all } \lambda \geq \delta. \tag{8.5}$$

Thus, for any $d^\lambda \in D^*_\lambda$ with $\lambda \geq \delta$, we have

$$G(r,d^\lambda) - \lambda G(c,d^\lambda) = G(b^\lambda,d^\lambda) \geq G(b^\lambda,d') = G(r,d') - \lambda G(c,d').$$

That is,

$$\frac{G(r,d^\lambda) - G(r,d')}{\lambda} + G(c,d') - \rho \geq G(c,d^\lambda) - \rho,$$

which, together with Assumption 8.2.1(ii) and (8.5), yields

$$G(c,d^\lambda) - \rho \leq \frac{|G(r,d^\lambda)| + |G(r,d')|}{\lambda} - \kappa \leq \frac{2M}{\lambda} - \kappa < 0 \text{ for all } \lambda \geq \delta. \tag{8.6}$$

Hence, it follows from (8.6) that $\widetilde{\lambda} \leq \delta < \infty$. □

Lemma 8.2.3. *Suppose that Assumptions 8.2.1(i) and (iii) hold. If* $\lim_{k\to\infty} \lambda_k = \lambda$, *and* $d^{\lambda_k} \in D^*_{\lambda_k}$ *(for each* $k \geq 1$*) is such that* $\lim_{k\to\infty} d^{\lambda_k} = \overline{d} \in D'$, *then* $\overline{d} \in D^*_\lambda$.

Proof. As $d^{\lambda_k} \in D^*_{\lambda_k}$ for all $k \geq 1$, by (8.1) and (8.3), we have

$$G(r,d^{\lambda_k}) - \lambda_k G(c,d^{\lambda_k}) = G(b^{\lambda_k},d^{\lambda_k}) \geq G(b^{\lambda_k},d) = G(r,d) - \lambda_k G(c,d) \tag{8.7}$$

for all $d \in D$. Letting $k \to \infty$ in (8.7) and using Assumption 8.2.1(iii), we get

$$G(b^\lambda,\overline{d}) = G(r,\overline{d}) - \lambda G(c,\overline{d}) \geq G(r,d) - \lambda G(c,d) = G(b^\lambda,d) \text{ for all } d \in D.$$

Thus, $\overline{d} \in D^*_\lambda$. □

Lemma 8.2.4. *If there exist* $\lambda_0 \geq 0$ *and* $d^* \in U$ *such that*

$$G(c,d^*) = \rho \quad \text{and} \quad G(b^{\lambda_0},d^*) = G^*(b^{\lambda_0}),$$

then d^* *is a constrained-optimal solution to the problem (8.2).*

Proof. For any $d \in U$, since $G(b^{\lambda_0},d^*) = G^*(b^{\lambda_0}) \geq G(b^{\lambda_0},d)$, we have

$$G(r,d^*) - \lambda_0 G(c,d^*) \geq G(r,d) - \lambda_0 G(c,d). \tag{8.8}$$

As $G(c,d^*) = \rho$ and $G(c,d) \leq \rho$ (because $d \in U$), from (8.8) we get

$$G(r,d^*) \geq G(r,d^*) + \lambda_0(G(c,d) - \rho) \geq G(r,d) \text{ for all } d \in U,$$

which implies the desired result. □

8.3 Main Results

In this section, we focus on the existence of a constrained-optimal solution. To do so, in addition to Assumptions 8.2.1 and 8.2.2, we also impose the following condition.

Assumption 8.3.1

(i) *For each* $\theta \in [0,1]$, $d_1, d_2 \in D^*_{\widetilde{\lambda}}$, $d_\theta := \theta d_1 + (1-\theta)d_2$ *satisfies* $G(b^{\widetilde{\lambda}}, d_\theta) = G^*(b^{\widetilde{\lambda}})$.

(ii) $G(c, d_\theta)$ *is continuous in* $\theta \in [0,1]$.

Remark 8.3.4. For each fixed $c_1 \in C$, if $G(c_1, \cdot)$ satisfies the following property

$$G(c_1, d_\theta) = \theta G(c_1, d_1) + (1-\theta)G(c_1, d_2)$$

for all $d_1, d_2 \in D$ and $\theta \in [0,1]$, then Assumption 8.3.1 is obviously true.

Now we give our first main result on the problem (8.2).

Theorem 8.3.1. *Under Assumptions 8.2.1, 8.2.2, and 8.3.1, the following statements hold:*

(a) *If* $\widetilde{\lambda} = 0$, *then there exists a constrained-optimal solution* $\widetilde{d} \in D'$.

(b) *If* $\widetilde{\lambda} > 0$, *then a constrained-optimal solution* $d^* \in D$ *exists, and moreover, there exist a number* $\theta^* \in [0,1]$ *and* $d^1, d^2 \in D^*_{\widetilde{\lambda}}$ *such that*

$$G(c, d^1) \geq \rho, \quad G(c, d^2) \leq \rho, \quad \text{and} \quad d^* = \theta^* d^1 + (1-\theta^*)d^2.$$

Proof. (a) The case $\widetilde{\lambda} = 0$: By the definition of $\widetilde{\lambda}$, there exists a sequence $d^{\lambda_k} \in D^*_{\lambda_k} \subset D'$ such that $\lambda_k \downarrow 0$ as $k \to \infty$. Because D' is compact, without loss of generality, we may assume that $d^{\lambda_k} \to \widetilde{d} \in D'$. Thus, by Lemma 8.2.1, we have $G(c, d^{\lambda_k}) \leq \rho$ for all $k \geq 1$, and then it follows from Assumption 8.2.1(iii) that $\widetilde{d} \in U$. Moreover, for each $d \in U$, we have $G^*(b^{\lambda_k}) = G(b^{\lambda_k}, d^{\lambda_k}) \geq G(b^{\lambda_k}, d)$, which, together with Assumption 8.2.1(ii), implies

$$G(r, d^{\lambda_k}) - G(r, d) \geq \lambda_k (G(c, d^{\lambda_k}) - G(c, d)) \geq -2\lambda_k M. \tag{8.9}$$

Letting $k \to \infty$ in (8.9), by Assumption 8.2.1(iii), we have

$$G(r, \widetilde{d}) \geq G(r, d) \quad \text{for all } d \in U,$$

which means that $\widetilde{d}$ is a constrained-optimal solution.

(b) The case $\widetilde{\lambda} \in (0, \infty)$: Since $\widetilde{\lambda}$ is in $(0, \infty)$, there exist two sequences of positive numbers $\{\lambda_k\}$ and $\{\delta_k\}$ such that $d^{\lambda_k} \in D^*_{\lambda_k}$, $d^{\delta_k} \in D^*_{\delta_k}$, $\lambda_k \uparrow \widetilde{\lambda}$, and $\delta_k \downarrow \widetilde{\lambda}$

as $k \to \infty$. By the compactness of D', we may suppose that $d^{\lambda_k} \to d^1 \in D'$ and $d^{\delta_k} \to d^2 \in D'$. By Lemma 8.2.3, we have $d^1, d^2 \in D^*_{\tilde{\lambda}}$. By Assumption 8.2.1(iii) and Lemma 8.2.1, we have

$$G(c, d^1) \geq \rho \ \text{ and } \ G(c, d^2) \leq \rho. \tag{8.10}$$

Define the following map:

$$\theta \mapsto G(c, \theta d^1 + (1-\theta)d^2) \ \text{ for each } \theta \in [0,1].$$

Thus, it follows from Assumption 8.3.1(ii) and (8.10) that there exists $\theta^* \in [0,1]$ such that

$$G(c, \theta^* d^1 + (1-\theta^*)d^2) = \rho. \tag{8.11}$$

Let $d^* := \theta^* d^1 + (1-\theta^*)d^2$. Then, by Assumption 8.3.1(i), we have $G(b^{\tilde{\lambda}}, d^*) = G^*(b^{\tilde{\lambda}})$, which together with (8.11) and Lemma 8.2.4 yields that $d^* \in D$ is a constrained-optimal solution. □

To further characterize a constrained-optimal solution, we next consider a particular case of the problem (8.2).

A special case: Let $X := \{1, 2, \ldots\}$, Y be a metric space, and $\mathscr{P}(Y)$ the set of all probability measures on Y. For each $i \in X$, $Y(i) \subset Y$ is assumed to be a compact metric space. Let $D := \{\psi |\ \psi : X \to \mathscr{P}(Y)$ such that $\psi(\cdot|i) \in \mathscr{P}(Y(i)) \ \ \forall\, i \in X\}$, and $D' := \{d |\ d : X \to Y$ such that $d(i) \in Y(i) \ \ \forall\, i \in X\}$.

Remark 8.3.5. (a) The set D is convex. That is, if $\psi_k (k = 1, 2)$ are in D, and $\psi^p(\cdot|i) := p\psi_1(\cdot|i) + (1-p)\psi_2(\cdot|i)$ for any $p \in [0,1]$ and $i \in X$, then $\psi^p \in D$.

(b) A function $d \in D'$ may be identified with the element $\psi \in D$, for which $\psi(i)$ is the Dirac measure at the point $d(i)$ for all $i \in X$. Hence, we have $D' \subset D$.

(c) Note that D' can be written as the product space $D' = \prod_{i \in X} Y(i)$. Hence, by the compactness of $Y(i)$ and the Tychonoff theorem, D' is a compact metric space.

In order to obtain the characterization of a constrained-optimal solution for this particular case, we also need the following condition.

Assumption 8.3.2 *For each $\lambda \geq 0$, if $d^1, d^2 \in D^*_\lambda$, then $d \in D^*_\lambda$ for each $d \in \{d \in D' : d(i) \in \{d^1(i), d^2(i)\}\ \forall\, i \in X\}$.*

Then, we have the second main result on the problem (8.2) as follows.

Theorem 8.3.2. *(For the special case.) Suppose that Assumptions 8.2.1, 8.2.2, 8.3.1, and 8.3.2 hold for the special case. Then there exists a constrained-optimal solution d^*, which is of one of the following two forms (i) and (ii): (i) $d^* \in D'$ and (ii) there exist $g^1, g^2 \in D^*_{\tilde{\lambda}}$, a point $i^* \in X$, and a number $\theta_0 \in [0,1]$ such that $g^1(i) = g^2(i)$ for all $i \neq i^*$, and, in addition,*

$$d^*(y|i) = \begin{cases} \theta_0 & for\ y = g^1(i)\ \ when\ \ i = i^*, \\ 1-\theta_0 & for\ y = g^2(i)\ \ when\ \ i = i^*, \\ 1 & for\ y = g^1(i)\ \ when\ \ i \neq i^*. \end{cases}$$

Proof. Let $\widetilde{\lambda}$ be as in (8.4). If $\widetilde{\lambda} = 0$, by Theorem 8.3.1 we have $d^* \in D'$. Thus, we only need to consider the other case $\widetilde{\lambda} > 0$. By Theorem 8.3.1(b), there exist $d^1, d^2 \in D^*_{\widetilde{\lambda}}$ such that $G(c,d^1) \geq \rho$ and $G(c,d^2) \leq \rho$. If $G(c,d^1)$ (or $G(c,d^2)$)$= \rho$, it follows from Lemma 8.2.4 that d^1 (or d^2) is a constrained-optimal solution. Hence, to complete the proof, we shall consider the following case:

$$G(c,d^1) > \rho \quad \text{and} \quad G(c,d^2) < \rho. \tag{8.12}$$

Using d^1 and d^2, we construct a sequence $\{d_n\}$ as follows. For all $n \geq 1$ and $i \in X$, let

$$d_n(i) = \begin{cases} d^1(i) & i < n, \\ d^2(i) & i \geq n. \end{cases}$$

Obviously, $d_1 = d^2$ and $\lim\limits_{n\to\infty} d_n = d^1$. Since $d^1, d^2 \in D^*_{\widetilde{\lambda}}$, by Assumption 8.3.2, we see that $d_n \in D^*_{\widetilde{\lambda}}$ for all $n \geq 1$. As $d_1 = d^2$, by (8.12) we have $G(c,d_1) < \rho$. If there exists n^* such that $G(c,d_{n^*}) = \rho$, then d_{n^*} is a constrained-optimal solution (by Lemma 8.2.4). Thus, in the remainder of the proof, we may assume that $G(c,d_n) \neq \rho$ for all $n \geq 1$. If $G(c,d_n) < \rho$ for all $n \geq 1$, then by Assumption 8.2.1(iii), we have

$$\lim_{n\to\infty} G(c,d_n) = G(c,d^1) \leq \rho,$$

which is a contradiction to (8.12). Hence, there exists some $n > 1$ such that $G(c,d_n) > \rho$, which, together with $G(c,d_1) < \rho$, gives the existence of some $\widetilde{n}$ such that

$$G(c,d_{\widetilde{n}}) < \rho \ \text{ and } \ G(c,d_{\widetilde{n}+1}) > \rho. \tag{8.13}$$

Obviously, $d_{\widetilde{n}}$ and $d_{\widetilde{n}+1}$ differ in at most the point $\widetilde{n}$.

Let $g^1 := d_{\widetilde{n}}$, $g^2 := d_{\widetilde{n}+1}$ and $i^* := \widetilde{n}$. For any $\theta \in [0,1]$, using g^1 and g^2, we construct $d_\theta \in D$ as follows. For each $i \in X$,

$$d_\theta(y|i) = \begin{cases} \theta & \text{for } \ y = g^1(i) \ \text{ when } \ i = i^*, \\ 1-\theta & \text{for } \ y = g^2(i) \ \text{ when } \ i = i^*, \\ 1 & \text{for } \ y = g^1(i) \ \text{ when } \ i \neq i^*. \end{cases}$$

Then, we have

$$d_\theta(\cdot|i) = \theta\delta_{g^1(i)}(\cdot) + (1-\theta)\delta_{g^2(i)}(\cdot) \tag{8.14}$$

for all $i \in X$ and $\theta \in [0,1]$, where $\delta_y(\cdot)$ denotes the Dirac measure at any point y. Hence, by (8.13), (8.14), and Assumption 8.3.1, there exists $\theta_0 \in (0,1)$ such that

$$G(c, d_{\theta_0}) = \rho \text{ and } G(b^{\widetilde{\lambda}}, d_{\theta_0}) = G^*(b^{\widetilde{\lambda}}),$$

which, together with Lemma 8.2.4, yield that d_{θ_0} is a constrained-optimal solution. Obviously, d_{θ_0} randomizes between g^1 and g^2, which differ in at most the point i^*, and so the theorem follows. □

8.4 Applications to MDPs with a Constraint

In this section, we show applications of the constrained optimization problem to MDPs with a constraint. In Sect. 8.4.1, we use Theorem 8.3.1 to show the existence of a constrained-optimal policy for the constrained discounted discrete-time MDPs in a Polish space and with unbounded rewards/costs. In Sect. 8.4.2, we investigate an application of Theorem 8.3.2 to discrete-time constrained MDPs with state-dependent discount factors. In Sects. 8.4.3 and 8.4.4, we will improve the corresponding results in [18, 20, 35] using Theorem 8.3.2 above.

8.4.1 Discrete-Time Constrained MDPs with Discounted Criteria

The constrained discounted discrete-time MDPs with a constant discount factor have been studied; see, for instance, [2, 11, 32] for the case of a countable state space and [15, 16, 24, 27, 29] for the case of a Borel state space. Except [2] dealing with the case in which the rewards may be unbounded from above and from below, all the aforementioned works investigate the case in which rewards are assumed to be bounded from above. To the best of our knowledge, in this subsection we first deal with the case in which the state space is a Polish space and the rewards may be unbounded from above and from below.

The model of discrete-time constrained MDPs under consideration is as follows [22, 23]:

$$\{X, A, (A(x), x \in X), Q(\cdot|x,a), r(x,a), c(x,a), \rho\}, \tag{8.15}$$

where X and A are state and action spaces, which are assumed to be Polish spaces with Borel σ-algebras $\mathscr{B}(X)$ and $\mathscr{B}(A)$, respectively. We denote by $A(x) \in \mathscr{B}(A)$ the set of admissible actions at state $x \in X$. Let $K := \{(x,a)|x \in X, a \in A(x)\}$, which is assumed to be a closed subset of $X \times A$. Furthermore, the transition law $Q(\cdot|x,a)$ with $(x,a) \in K$ is a stochastic kernel on X given K. Finally, the function $r(x,a)$ on

K denotes rewards, while the function $c(x,a)$ on K and the number ρ denote costs and a constraint, respectively. We assume that $r(x,a)$ and $c(x,a)$ are real-valued Borel-measurable on K.

We denote by Π, Φ, and F the classes of all randomized history-dependent policies, randomized stationary policies, and stationary policies, respectively; see [22, 23] for details.

Let $\Omega := (X \times A)^\infty$ and $\mathscr{F}$ the corresponding product σ-algebra. Then, for an arbitrary policy $\pi \in \Pi$ and an arbitrary initial distribution ν on X, the well-known Tulcea theorem [22, p.178] gives the existence of a unique probability measure P_ν^π on $(\Omega, \mathscr{F})$ and a stochastic process $\{(x_k, a_k), k \geq 0\}$. The expectation operator with respect to P_ν^π is denoted by E_ν^π, and we write E_ν^π as E_x^π when $\nu(\{x\}) = 1$.

Fix a discount factor $\alpha \in (0,1)$ and an initial distribution ν on X. We define the expected discounted reward $V(r,\pi)$ and the expected discounted cost $V(c,\pi)$ as follows:

$$V(r,\pi) := E_\nu^\pi\left[\sum_{k=0}^{\infty} \alpha^k r(x_k,a_k)\right] \text{ and } V(c,\pi) := E_\nu^\pi\left[\sum_{k=0}^{\infty} \alpha^k c(x_k,a_k)\right] \text{ for all } \pi \in \Pi.$$

Then, the constrained optimization problem for the model (8.15) is as follows:

$$\text{Maximize } V(r,\cdot) \text{ over } U_1 := \big\{\pi \in \Pi \,|\, V(c,\pi) \leq \rho\big\}. \tag{8.16}$$

To solve (8.16), we consider the following conditions:

(B1) There exist a continuous function $\omega_1 \geq 1$ on X and positive constants L_1, m, and $\beta_1 < 1$ such that, for each $(x,a) \in K$,

$$|r(x,a)| \leq L_1\omega_1(x),\ |c(x,a)| \leq L_1\omega_1(x), \text{ and } \int_X \omega_1^2(y)Q(\mathrm{d}y|x,a) \leq \beta_1\omega_1^2(x) + m.$$

(B2) The function ω_1 is a moment function on K, that is, there exists a nondecreasing sequence of compact sets $K_n \uparrow K$ such that $\liminf_{n\to\infty}\big\{\omega_1(x) : (x,a) \notin K_n\big\} = \infty$.

(B3) $\nu(\omega_1^2) := \int_X \omega_1^2(x)\nu(\mathrm{d}x) < \infty$.

(B4) $Q(\cdot|x,a)$ is weakly continuous on K, that is, the function $\int_X u(y)Q(\mathrm{d}y|x,a)$ is continuous in $(x,a) \in K$ for each bounded continuous function u on X.

(B5) The functions $r(x,a)$ and $c(x,a)$ are continuous on K.

(B6) There exists $\pi' \in \Pi$ such that $V(c,\pi') < \rho$.

Remark 8.4.6. (a) Condition (B1) is known as the statement of the Lyapunov-like inequality and the growth condition on the rewards/costs. Conditions (B4) and (B5) are the usual continuity conditions. Condition (B6) is the Slater-like condition.

(b) Conditions (B1) and (B3) are used to guarantee the finiteness of the expected discounted rewards/costs. The role of condition (B2) is to prove the compactness of the set of all the discount occupation measures in the ω_1-weak topology (see Lemma 8.4.5).

To state our main results of Sect. 8.4.1, we need to introduce some notation. Let ω_1 be as in condition (B1). We denote by $B_{\omega_1}(X)$ the Banach space of real-valued measurable functions u on X with the finite norm $\|u\|_{\omega_1} := \sup_{x\in X} \frac{|u(x)|}{\omega_1(x)}$, that is, $B_{\omega_1}(X) := \{u|\ \|u\|_{\omega_1} < \infty\}$. Moreover, we say that a function v on K belongs to $B_{\omega_1}(K)$ if $x \mapsto \sup_{a\in A(x)} |v(x,a)|$ is in $B_{\omega_1}(X)$. We denote by $C_{\omega_1}(K)$ the set of all continuous functions on K which also belong to $B_{\omega_1}(K)$, and $\mathscr{M}_{\omega_1}(K)$ stands for the set of all measures μ on $\mathscr{B}(K)$ such that $\int_K \omega_1(x)\mu(\mathrm{d}x,\mathrm{d}a) < \infty$. Moreover, $\mathscr{M}_{\omega_1}(K)$ is endowed with ω_1-weak topology. Recall that the ω_1-weak topology on $\mathscr{M}_{\omega_1}(K)$ is the coarsest topology for which all mappings $\mu \mapsto \int_K v(x,a)\mu(\mathrm{d}x,\mathrm{d}a)$ are continuous for each $v \in C_{\omega_1}(K)$. Since X and A are both Polish spaces, by Corollary A.44 in [14, p.423] we see that $\mathscr{M}_{\omega_1}(K)$ is metrizable with respect to the ω_1-weak topology.

By Lemma 24 in [29, p.141], it suffices to consider the discount occupation measures induced by randomized stationary policies in Φ. For each $\varphi \in \Phi$, we define the discount occupation measure by

$$\eta^{\varphi}(B\times E) := \sum_{k=0}^{\infty} \alpha^k P_v^{\varphi}(x_k \in B, a_k \in E) \quad \text{for all } B \in \mathscr{B}(X) \text{ and } E \in \mathscr{B}(A).$$

The set of all the discount occupation measures is denoted by $\mathscr{N}$, i.e., $\mathscr{N} := \{\eta^{\varphi} : \varphi \in \Phi\}$. From the conditions (B1) and (B3), we have

$$\int_K \omega_1(x)\eta^{\varphi}(\mathrm{d}x,\mathrm{d}a) = \sum_{k=0}^{\infty} \alpha^k E_v^{\varphi}\left[\omega_1(x_k)\right] \le \frac{\nu(\omega_1^2)}{1-\alpha} + \frac{m}{(1-\beta_1)(1-\alpha)} < \infty \tag{8.17}$$

for all $\varphi \in \Phi$, which yields $\mathscr{N} \subset \mathscr{M}_{\omega_1}(K)$.

Then, the constrained optimization problem (8.16) is equivalent to the following form:

$$\text{Maximize} \int_K r(x,a)\eta(\mathrm{d}x,\mathrm{d}a) \text{ over } \left\{\eta \in \mathscr{N} | \int_K c(x,a)\eta(\mathrm{d}x,\mathrm{d}a) \le \rho\right\} =: U_o. \tag{8.18}$$

Now we provide a characterization of the discount occupation measures below.

Lemma 8.4.5. *Under conditions (B1)–(B4), the following statements hold:*

(a) If $\eta \in \mathscr{M}_{\omega_1}(K)$, then η is in $\mathscr{N}$ if and only if

$$\int_K u(x)\eta(\mathrm{d}x,\mathrm{d}a) = \int_X u(x)\nu(\mathrm{d}x) + \alpha \int_K \int_X u(y)Q(\mathrm{d}y|x,a)\eta(\mathrm{d}x,\mathrm{d}a)$$

for each bounded continuous function u on X.

(b) $\mathscr{N}$ is convex and compact in the ω_1-weak topology.

Proof. (a) See Lemma 25 in [29, p.141] for the proof of part (a).

(b) The convexity property follows directly from part (a). To prove that $\mathscr{N}$ is compact, we will first show that $\mathscr{N}$ is closed in the ω_1-weak topology. Let $\{\eta^{\varphi_n}\} \subset \mathscr{N}$ be a sequence converging to some measure η on $X \times A$ in the ω_1-weak topology. Thus, there exists a positive integer N_1 such that for each $n \geq N_1$, we have

$$\left| \int_K \omega_1(x)\eta^{\varphi_n}(\mathrm{d}x,\mathrm{d}a) - \int_{X\times A} \omega_1(x)\eta(\mathrm{d}x,\mathrm{d}a) \right| \leq 1,$$

which together with (8.17) yields

$$\int_{X\times A} \omega_1(x)\eta(\mathrm{d}x,\mathrm{d}a) \leq \frac{\nu(\omega_1^2)}{1-\alpha} + \frac{m}{(1-\beta_1)(1-\alpha)} + 1 < \infty,$$

and so $\eta \in \mathscr{M}_{\omega_1}(X \times A)$. Moreover, since K is assumed to be closed and η^{φ_n} weakly converges to η, by Theorem A.38 in [14, p.420] we have

$$0 = \liminf_{n\to\infty} \eta^{\varphi_n}(K^c) \geq \eta(K^c) \geq 0,$$

which implies that η concentrates on K, where K^c denotes the complement of K. In addition, by part (a) we have

$$\int_K u(x)\eta^{\varphi_n}(\mathrm{d}x,\mathrm{d}a) = \int_X u(x)\nu(\mathrm{d}x) + \alpha \int_K \int_X u(y)Q(\mathrm{d}y|x,a)\eta^{\varphi_n}(\mathrm{d}x,\mathrm{d}a)$$

for each bounded continuous function u on X, which together with condition (B4) yields

$$\int_K u(x)\eta(\mathrm{d}x,\mathrm{d}a) = \int_X u(x)\nu(\mathrm{d}x) + \alpha \int_K \int_X u(y)Q(\mathrm{d}y|x,a)\eta(\mathrm{d}x,\mathrm{d}a).$$

Hence, by part (a) we see that $\eta \in \mathscr{N}$, and so $\mathscr{N}$ is closed.

To prove the compactness of $\mathscr{N}$, it suffices to show that $\mathscr{N}$ is relatively compact in the ω_1-weak topology. By (8.17) we have

$$\begin{aligned} \sup_{\eta\in\mathscr{N}} \int_K \omega_1(x)\eta(\mathrm{d}x,\mathrm{d}a) &= \sup_{\varphi\in\Phi} \int_K \omega_1(x)\eta^{\varphi}(\mathrm{d}x,\mathrm{d}a) \\ &\leq \frac{\nu(\omega_1^2)}{1-\alpha} + \frac{m}{(1-\beta_1)(1-\alpha)} < \infty. \end{aligned} \tag{8.19}$$

On the other hand, from condition (B2), we see that $\varpi_n := \inf\{\omega_1(x) : (x,a) \notin K_n\} \uparrow \infty$. Then, by conditions (B1) and (B3), we have

$$\begin{aligned} \varpi_n \int_{K_n^c} \omega_1(x)\eta^{\varphi}(\mathrm{d}x,\mathrm{d}a) &\leq \int_K \omega_1^2(x)\eta^{\varphi}(\mathrm{d}x,\mathrm{d}a) \\ &\leq \frac{\nu(\omega_1^2)}{1-\alpha} + \frac{m}{(1-\beta_1)(1-\alpha)} \end{aligned} \tag{8.20}$$

for all $\varphi \in \Phi$. Thus, by (8.20) we see that for any $\varepsilon > 0$, there exists an integer $N_2 > 0$ such that

$$\sup_{\varphi\in\Phi} \int_{K^c_{N_2}} \omega_1(x)\eta^{\varphi}(\mathrm{d}x,\mathrm{d}a) \le \varepsilon. \tag{8.21}$$

Hence, by (8.19), (8.21), and Corollary A.46 in [14, p.424], we conclude that $\mathscr{N}$ is relatively compact in the ω_1-weak topology. Therefore, $\mathscr{N}$ is compact in the ω_1-weak topology. □

Under conditions (B1)–(B5), from Lemma 8.4.5 and (8.18), we define a real-valued function G on $C \times D := C_{\omega_1}(K) \times \mathscr{N}$ as follows:

$$G(c,\eta) := \int_K c(x,a)\eta(\mathrm{d}x,\mathrm{d}a) \quad \text{for } (c,\eta) \in C \times D = C_{\omega_1}(K) \times \mathscr{N}. \tag{8.22}$$

Obviously, the function G defined in (8.22) satisfies (8.1). Moreover, let $D' := \mathscr{N}$.

Now we provide our main result of Sect. 8.4.1 on the existence of constrained-optimal policies for (8.16).

Proposition 8.4.1. *Under conditions (B1)–(B6), there exists a constrained-optimal policy $\varphi^* \in \Phi$ for the constrained MDPs in (8.16), that is, $V(r,\varphi^*) = \sup_{\pi\in U_1} V(r,\pi)$.*

Proof. We first verify Assumption 8.2.1. From conditions (B1) and (B5), we see that for each $\lambda \ge 0$, the mapping $\eta^{\varphi} \mapsto \int_K (r(x,a) - \lambda c(x,a))\eta^{\varphi}(\mathrm{d}x,\mathrm{d}a)$ is continuous on $\mathscr{N}$. Thus, Assumption 8.2.1(i) follows from the compactness of $\mathscr{N}$. Moreover, by condition (B1) and (8.17), we have

$$\max\{|G(r,\eta)|, |G(c,\eta)|\} \le \frac{L_1 \nu(\omega_1^2)}{1-\alpha} + \frac{mL_1}{(1-\beta_1)(1-\alpha)} =: M$$

for all $\eta \in \mathscr{N}$, and so Assumption 8.2.1(ii) follows. By conditions (B1) and (B5), we see that Assumption 8.2.1(iii) is obviously true.

Secondly, Assumption 8.2.2 follows from condition (B6) and Lemma 24 in [29, p.141].

Finally, since $G(c,\theta\eta_1 + (1-\theta)\eta_2) = \theta G(c,\eta_1) + (1-\theta)G(c,\eta_2)$ for all $\theta \in [0,1]$ and $\eta_1, \eta_2 \in \mathscr{N}$, Assumption 8.3.1 is obviously true.

Hence, Theorem 8.3.1 gives the existence of $\eta^* \in \mathscr{N}$ such that

$$\int_K r(x,a)\eta^*(\mathrm{d}x,\mathrm{d}a) = \sup_{\eta\in U_o} \int_K r(x,a)\eta(\mathrm{d}x,\mathrm{d}a),$$

which together with Theorem 6.3.7 in [22] implies Proposition 8.4.1. □

8.4.2 Constrained MDPs with State-Dependent Discount Factors

In this subsection, we use discrete-time constrained MDPs with state-dependent discount factors to present another application of the constrained optimization problem. Discrete-time unconstrained MDPs with nonconstant discount factors are studied in [26, 34]. Moreover, [36] deals with discrete-time constrained MDPs with state-dependent discount factors in which the costs are assumed to be bounded from below by a convex analytic approach, and here we use Theorem 8.3.1 above to deal with the case in which the costs are allowed to be unbounded from above and from below.

The model of discrete-time constrained MDPs with state-dependent discount factors is as follows:

$$\big\{X, A, (A(i), i \in X), Q(\cdot|i,a), (\alpha(i), i \in X), r(i,a), c(i,a), \rho\big\},$$

where the state space X is the set of all positive integers, $\alpha(i) \in (0,1)$ are given discount factors depending on state $i \in X$, and the other components are the same as in (8.15), with i_k here in lieu of x_k in Sect. 8.4.1.

Fix any initial distribution ν on X. The discounted criteria, $W(r,\pi)$ and $W(c,\pi)$, are defined by

$$W(r,\pi) := E_\nu^\pi\left[r(i_0,a_0) + \sum_{n=1}^{\infty}\prod_{k=0}^{n-1}\alpha(i_k)r(i_n,a_n)\right],$$

$$W(c,\pi) := E_\nu^\pi\left[c(i_0,a_0) + \sum_{n=1}^{\infty}\prod_{k=0}^{n-1}\alpha(i_k)c(i_n,a_n)\right] \text{ for all } \pi \in \Pi.$$

Then, the constrained optimization problem is as follows:

$$\sup_{\pi\in\Pi} W(r,\pi) \text{ subject to } W(c,\pi) \le \rho. \tag{8.23}$$

To ensure the existence of a constrained-optimal policy π^* for (8.23) (i.e., $W(r,\pi^*) \ge W(r,\pi)$ for all π such that $W(c,\pi) \le \rho$), we consider the following conditions from [34]:

(C1) There exists a constant $\overline{\alpha} \in (0,1)$ such that $0 < \alpha(i) \le \overline{\alpha}$ for all $i \in X$.

(C2) There exist constants $L_2 > 0$ and β_2, with $1 \le \beta_2 < \frac{1}{\overline{\alpha}}$ and a function $\omega_2 \ge 1$ on X such that, for each $(i,a) \in K$,

$$|r(i,a)| \le L_2\omega_2(i),\ |c(i,a)| \le L_2\omega_2(i), \text{ and } \sum_{j\in X}\omega_2(j)Q(j|i,a) \le \beta_2\omega_2(i).$$

(C3) $\nu(\omega_2) := \sum_{i\in X}\omega_2(i)\nu(i) < \infty$.

(C4) For each $i \in X$, $A(i)$ is compact.
(C5) For each $i, j \in X$, the functions $r(i,a)$, $c(i,a)$, $Q(j|i,a)$, and $\sum_{k \in X} \omega_2(k) Q(k|i,a)$ are continuous in $a \in A(i)$.
(C6) There exists $\widetilde{\pi} \in \Pi$ such that $W(c, \widetilde{\pi}) < \rho$.

Remark 8.4.7. Conditions (C1)–(C3) are known as the finiteness conditions. Conditions (C4) and (C5) are the usual continuity-compactness conditions. Condition (C6) is the Slater-like condition.

Under conditions (C1)–(C5), we define a real-valued function G on $C \times D := C_{\omega_2}(K) \times \Pi$ as follows:

$$G(c,\pi) := W(c,\pi) \quad \text{for } (c,\pi) \in C \times D = C_{\omega_2}(K) \times \Pi. \tag{8.24}$$

Obviously, since the set Π is convex, the function G defined in (8.24) satisfies (8.1).

Let $D' := F$. Then, we state our main result of Sect. 8.4.2 on the existence of constrained-optimal policies for (8.23).

Theorem 8.4.3. *Suppose that conditions (C1)–(C6) hold. Then there exists a constrained-optimal policy for (8.23), which is either a stationary policy or a randomized stationary policy that randomizes between two stationary policies differing in at most one state; that is, there exist two stationary policies f^1, f^2, a state $i^* \in X$, and a number $p^* \in [0,1]$ such that $f^1(i) = f^2(i)$ for all $i \neq i^*$, and, in addition, the randomized stationary policy $\pi^{p^*}(\cdot|i)$ is constrained-optimal, where*

$$\pi^{p^*}(a|i) = \begin{cases} p^* & for\ a = f^1(i) \ \ when \ \ i = i^*, \\ 1 - p^* & for\ a = f^2(i) \ \ when \ \ i = i^*, \\ 1 & for\ a = f^1(i) \ \ when \ \ i \neq i^*. \end{cases}$$

Remark 8.4.8. Theorem 8.4.3 extends the corresponding one in [32] for a constant discount factor to the case of state-dependent discount factors. Moreover, we remove the nonnegativity assumption on the costs as in [32].

We will prove Theorem 8.4.3 using Theorem 8.3.2. To do so, we introduce the notation below.

For each $\varphi \in \Phi$, $i \in X$, and $\pi \in \Pi$, define

$$W_r(i,\pi) := E_i^{\pi}\left[r(i_0,a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \alpha(i_k) r(i_n,a_n) \right],$$

$$W_c(i,\pi) := E_i^{\pi}\left[c(i_0,a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \alpha(i_k) c(i_n,a_n) \right],$$

$$b^{\lambda}(i,a) := r(i,a) - \lambda c(i,a) \ \text{ for all } (i,a) \in K,$$

$$W^{\lambda}(i,\pi) := E_i^{\pi}\left[b^{\lambda}(i_0,a_0) + \sum_{n=1}^{\infty}\prod_{k=0}^{n-1}\alpha(i_k)b^{\lambda}(i_n,a_n)\right],$$

$$W_{\lambda}^{*}(i) := \sup_{\pi\in\Pi} W^{\lambda}(i,\pi),$$

and

$$u(i,\varphi) := \int_{A(i)} u(i,a)\varphi(\mathrm{d}a|i), \quad \text{for } u(i,a) = b^{\lambda}(i,a), r(i,a), c(i,a),$$

$$Q(j|i,\varphi) := \int_{A(i)} Q(j|i,a)\varphi(\mathrm{d}a|i) \ \text{ for } j \in X.$$

Then, we give three lemmas below, which are used to prove Theorem 8.4.3.

Lemma 8.4.6. *Under conditions (C1)–(C5), the following assertions hold:*

(a) $|W(r,\pi)| \le \frac{L_2 v(\omega_2)}{1-\overline{\alpha}\beta_2}$ *and* $|W(c,\pi)| \le \frac{L_2 v(\omega_2)}{1-\overline{\alpha}\beta_2}$ *for all* $\pi \in \Pi$.

(b) For each fixed $\varphi \in \Phi$, $W_u(\cdot,\varphi)$ $(u = r,c)$ *is the unique solution in* $B_{\omega_2}(X)$ *to the following equation:*

$$v(i) = u(i,\varphi) + \alpha(i)\sum_{j\in X} v(j)Q(j|i,\varphi) \quad \textit{for all } i \in X.$$

(c) $W(r,f)$ *and* $W(c,f)$ *are continuous in* $f \in F$.

Proof. For the proofs of (a) and (b), see Theorem 3.1 in [34].
(c) We only prove the continuity of $W(r,f)$ in $f \in F$ because the other case is similar. Let $f_n \to f$ as $n \to \infty$, and fix any $i \in X$. Choose any subsequence $\{W_r(i,f_{n_m})\}$ of $\{W_r(i,f_n)\}$ converging to some point $v(i)$ as $m \to \infty$. Then, since X is denumerable, the Tychonoff theorem, together with the part (a) and $f_n \to f$, gives the existence of subsequence $\{W_r(j,f_{n_k}), j \in X\}$ of $\{W_r(j,f_{n_m}), j \in X\}$ such that

$$\lim_{k\to\infty} W_r(j,f_{n_k}) =: v'(j),\ v'(i) = v(i), \text{ and } \lim_{k\to\infty} f_{n_k}(j) = f(j) \ \text{ for all } j \in X. \quad (8.25)$$

Furthermore, by Theorem 3.1 in [34], we have $|v'(j)| \le \frac{L_2\omega_2(j)}{1-\overline{\alpha}\beta_2}$ for all $j \in X$, which implies that $v' \in B_{\omega_2}(X)$. On the other hand, for the given $i \in X$ and all $k \ge 1$, by part (b) we have

$$W_r(i,f_{n_k}) = r(i,f_{n_k}) + \alpha(i)\sum_{j\in X} W_r(j,f_{n_k})Q(j|i,f_{n_k}). \quad (8.26)$$

Then, it follows from (8.25), (8.26), condition (C5), and Lemma 8.3.7 in [23, p.48] that

$$v'(i) = r(i,f) + \alpha(i)\sum_{j\in X} v'(j)Q(j|i,f).$$

Hence, part (b) yields

$$v'(i) = W_r(i,f). \tag{8.27}$$

Thus, as the above subsequence $\{W_r(i,f_{n_m})\}$ and $i \in X$ are arbitrarily chosen and (by (8.27)) all such subsequences have the same limit $W_r(i,f)$, we have

$$\lim_{n\to\infty} W_r(i,f_n) = W_r(i,f) \quad \text{for all } i \in X.$$

Therefore, from condition (C3) and Theorem A.6 in [22, p.171], we get

$$\lim_{n\to\infty} W(r,f_n) = \sum_{i\in X}\left[\lim_{n\to\infty} W_r(i,f_n)\right]\nu(i) = \sum_{i\in X} W_r(i,f)\nu(i) = W(r,f),$$

which gives the desired conclusion, $W(r,f_n) \to W(r,f)$ as $n \to \infty$. □

Lemma 8.4.7. *Suppose that conditions (C1), (C2), (C4), and (C5) hold. Then we have*

(a) W_λ^* is the unique solution of the following equation in $B_{\omega_2}(X)$:

$$v(i) = \sup_{a\in A(i)}\left\{b^\lambda(i,a) + \alpha(i)\sum_{j\in X} v(j)Q(j|i,a)\right\} \quad \text{for all } i \in X. \tag{8.28}$$

(b) There exists a function $f^* \in F$ such that $f^*(i) \in A(i)$ attains the maximum in (8.28) for each $i \in X$, that is,

$$W_\lambda^*(i) = b^\lambda(i,f^*) + \alpha(i)\sum_{j\in X} W_\lambda^*(j)Q(j|i,f^*) \quad \text{for all } i \in X, \tag{8.29}$$

and $f^* \in F$ is optimal. Conversely, if $f^* \in F$ is optimal, it satisfies (8.29).

Proof. See Theorem 3.2 in [34]. □

Remark 8.4.9. Under conditions (C1), (C2), (C4), and (C5), for each $\lambda \geq 0$, let $D_\lambda^*(e) := \{f \in F : W^\lambda(i,f) = W_\lambda^*(i) \text{ for all } i \in X\}$. Then, it follows from Lemma 8.4.7 that $D_\lambda^*(e) \neq \emptyset$, and that $f \in D_\lambda^*(e)$ if and only if $f \in F$ satisfies (8.29).

Lemma 8.4.8. *Suppose that conditions (C1)–(C5) hold. Then, for each $f_1, f_2 \in D_\lambda^*(e)$ (with any fixed $\lambda \geq 0$), and $0 \leq p \leq 1$, define a policy π^p by $\pi^p(\cdot|i) := p\delta_{f_1(i)}(\cdot) + (1-p)\delta_{f_2(i)}(\cdot)$ for all $i \in X$. Then,*

(a) $W^\lambda(i,\pi^p) = W_\lambda^*(i)$ for all $i \in X$.
(b) $W(c,\pi^p)$ is continuous in $p \in [0,1]$.

Proof. (a) Since

$$b^{\lambda}(i,\pi^{p}) = pb^{\lambda}(i,f_1) + (1-p)b^{\lambda}(i,f_2), \tag{8.30}$$

$$Q(j|i,\pi^{p}) = pQ(j|i,f_1) + (1-p)Q(j|i,f_2), \tag{8.31}$$

by Lemma 8.4.7 and the definition of $D^{*}_{\lambda}(e)$, we have

$$W^{*}_{\lambda}(i) = b^{\lambda}(i,f_l) + \alpha(i)\sum_{j\in X} W^{*}_{\lambda}(j)Q(j|i,f_l) \text{ for all } i \in X \text{ and } l = 1,2,$$

which together with (8.30) and (8.31) gives

$$W^{*}_{\lambda}(i) = b^{\lambda}(i,\pi^{p}) + \alpha(i)\sum_{j\in X} W^{*}_{\lambda}(j)Q(j|i,\pi^{p}) \text{ for all } i \in X. \tag{8.32}$$

Therefore, by Lemma 8.4.6(b) and (8.32), we have $W^{\lambda}(i,\pi^{p}) = W^{*}_{\lambda}(i)$ for all $i \in X$, and so part (a) follows.

(b) For any $p \in [0,1]$ and any sequence $\{p_m\}$ in $[0,1]$ such that $\lim_{m\to\infty} p_m = p$, by Lemma 8.4.6, we have

$$W_c(i,\pi^{p_m}) = c(i,\pi^{p_m}) + \alpha(i)\sum_{j\in X} W_c(j,\pi^{p_m})Q(j|i,\pi^{p_m}) \text{ for all } i \in X \text{ and } m \geq 1. \tag{8.33}$$

Hence, as in the proof of Lemma 8.4.6, from (8.33), the definition of π^{p_m} and Theorem A.6 in [22, p.171], we have

$$\lim_{m\to\infty} W(c,\pi^{p_m}) = W(c,\pi^{p}),$$

and so $W(c,\pi^{p})$ is continuous in $p \in [0,1]$. □

Proof of Theorem 8.4.3. By Lemmas 8.4.6–8.4.8 and (8.24), we see that Assumptions 8.2.1 and 8.3.1 hold. Moreover, Assumptions 8.2.2 and 8.3.2 follow from condition (C6) and Lemma 8.4.7, respectively. Hence, by Theorem 8.3.2, we complete the proof. □

8.4.3 *Continuous-Time Constrained MDPs with Average Criteria*

In this subsection, removing the nonnegativity assumption on the cost function as in [20, 35], we will prove that the corresponding results in [20, 35] still hold using Theorem 8.3.2 above.

The model of continuous-time constrained MDPs is of the form [18, 20, 30, 35]:

$$\{X, A, (A(i), i \in X), q(\cdot|i,a), r(i,a), c(i,a), \rho\}, \tag{8.34}$$

where X is assumed to be a denumerable set. Without loss of generality, we assume that X is the set of all positive integers. Furthermore, the transition rates $q(j|i,a)$, which satisfy $q(j|i,a) \geq 0$ for all $(i,a) \in K$ and $j \neq i$. We also assume that the transition rates $q(j|i,a)$ are conservative, i.e., $\sum_{j \in X} q(j|i,a) = 0$ for all $(i,a) \in K$, and stable, which means that $q^*(i) := \sup_{a \in A(i)} -q(i|i,a) < \infty$ for all $i \in X$. In addition, $q(j|i,a)$ is measurable in $a \in A(i)$ for each fixed $i, j \in X$. The other components are the same as in (8.15), with a state i here in lieu of a state x in Sect. 8.4.1.

We denote by Π_m, Φ and F the classes of all randomized Markov policies, randomized stationary policies, and stationary policies, respectively; see [18,20,30,35] for details.

To guarantee the regularity of the Q-process, we impose the following drift condition from [30]:

(D1) There exists a nondecreasing function $\omega_3 \geq 1$ on X such that $\lim_{i \to \infty} \omega_3(i) = \infty$.
(D2) There exist constants $\gamma_1 \geq \kappa_1 > 0$ and a state $i_0 \in X$ such that

$$\sum_{j \in X} q(j|i,a)\omega_3^2(j) \leq -\kappa_1 \omega_3^2(i) + \gamma_1 I_{i_0}(i) \quad \text{for all } (i,a) \in K,$$

where $I_B(\cdot)$ denotes the indicator function of any set B.
Let $\overline{T} := [0,\infty)$, and let $(\Omega, \mathscr{B}(\Omega))$ be the canonical product measurable space with $(X \times A)^{\overline{T}}$ being the set of all maps from $\overline{T}$ to $X \times A$. Fix an initial distribution ν on X. Then, under conditions (D1) and (D2), by Theorem 2.3 in [20, p.14], for each policy $\pi \in \Pi_m$, there exist a unique probability measure P_ν^π on $(\Omega, \mathscr{B}(\Omega))$ and a stochastic process $\{(x(t), a(t)), t \geq 0\}$. The expectation operator with respect to P_ν^π is denoted by E_ν^π.

For each $\pi \in \Pi_m$, we define the expected average criteria, $\overline{V}(r,\pi)$ and $\overline{V}(c,\pi)$, as follows:

$$\overline{V}(r,\pi) := \liminf_{T \to \infty} \frac{E_\nu^\pi \left[\int_0^T r(x(t), a(t)) \mathrm{d}t \right]}{T},$$

$$\overline{V}(c,\pi) := \limsup_{T \to \infty} \frac{E_\nu^\pi \left[\int_0^T c(x(t), a(t)) \mathrm{d}t \right]}{T}.$$

Then, the constrained optimization problem for the average criteria is as follows:

$$\sup_{\pi \in \Pi_m} \overline{V}(r,\pi) \text{ subject to } \overline{V}(c,\pi) \leq \rho. \tag{8.35}$$

To guarantee the existence of a constrained-optimal policy π^* for (8.35) (i.e., $\overline{V}(r,\pi^*) \geq \overline{V}(r,\pi)$ for all π such that $\overline{V}(c,\pi) \leq \rho$), we need the following conditions from [20,30,35]:

(D3) There exists a constant $L_3 > 0$ such that

$$|r(i,a)| \le L_3\omega_3(i) \ \text{ and } \ |c(i,a)| \le L_3\omega_3(i) \ \text{ for all } (i,a) \in K.$$

(D4) For each $i \in X$, $A(i)$ is compact.
(D5) For each $i \in X$, $q^*(i) \le \omega_3(i)$.
(D6) $\nu(\omega_3^2) := \sum_{i\in X} \omega_3^2(i)\nu(i) < \infty$.
(D7) For each $i, j \in X$, the functions $r(i,a)$, $c(i,a)$, $q(j|i,a)$, and $\sum_{k\in X} \omega_3(k)q(k|i,a)$ are continuous in $a \in A(i)$. (D8) For each $\varphi \in \Phi$, the corresponding Markov process with transition rates $q(j|i,\varphi)$ is irreducible, where $q(j|i,\varphi) := \int_{A(i)} q(j|i,a)\varphi(\mathrm{d}a|i)$ for all $i, j \in X$.
(D9) There exists $\varphi' \in \Phi$ such that $\overline{V}(c,\varphi') < \rho$.

From conditions (D1), (D2), and (D8), by Theorem 4.2 in [28], for each $\varphi \in \Phi$, the corresponding Markov process with transition rates $q(j|i,\varphi)$ has a unique invariant probability measure μ_φ on X. Moreover, under conditions (D1)–(D4), (D7), and (D8), by Theorem 7.2 in [30] we have

$$\overline{V}(r,\varphi) = \lim_{T\to\infty} \frac{1}{T} E_\nu^\varphi \left[\int_0^T r(x(t),a(t))\mathrm{d}t \right] = \sum_{i\in X} r(i,\varphi)\mu_\varphi(i) \tag{8.36}$$

and

$$\overline{V}(c,\varphi) = \lim_{T\to\infty} \frac{1}{T} E_\nu^\varphi \left[\int_0^T c(x(t),a(t))\mathrm{d}t \right] = \sum_{i\in X} c(i,\varphi)\mu_\varphi(i), \tag{8.37}$$

where

$$r(i,\varphi) := \int_{A(i)} r(i,a)\varphi(\mathrm{d}a|i) \ \text{ and } \ c(i,\varphi) := \int_{A(i)} c(i,a)\varphi(\mathrm{d}a|i) \ \text{ for all } i \in X.$$

For each $\varphi \in \Phi$, we define the average occupation measure $\widehat{\mu}_\varphi$ by

$$\widehat{\mu}_\varphi(\{i\} \times B) := \mu_\varphi(i)\varphi(B|i) \ \text{ for all } i \in X \text{ and } B \in \mathscr{B}(A(i)).$$

The set of all the average occupation measures is denoted by $\mathscr{N}_1$, i.e., $\mathscr{N}_1 := \{\widehat{\mu}_\varphi : \varphi \in \Phi\}$.

Then, we have the following result.

Lemma 8.4.9. *Suppose that conditions (D1)–(D8) hold. Then, for any $\pi \in \Pi_m$ with $\overline{V}(c,\pi) \le \rho$, there exists $\widehat{\mu}_{\varphi_0} \in \mathscr{N}_1$ such that $\overline{V}(r,\varphi_0) \ge \overline{V}(r,\pi)$ and $\overline{V}(c,\varphi_0) \le \rho$.*

Proof. For each fixed $\pi \in \Pi_m$ with $\overline{V}(c,\pi) \le \rho$ and $n \ge 1$, we define a measure μ_n by

$$\mu_n(\{i\} \times B) := \frac{1}{n} \int_0^n E_\nu^\pi \left[I_{\{i\}\times B}(x(t),a(t)) \right] \mathrm{d}t \ \text{ for all } i \in X \text{ and } B \in \mathscr{B}(A(i)). \tag{8.38}$$

Then, by conditions (D2) and (D6), Lemma 6.3 in [20, p.90], and (8.38), we have

$$\begin{aligned}\sum_{i\in X}\int_{A(i)}\omega_3^2(i)\mu_n(i,\mathrm{d}a) &= \frac{1}{n}\int_0^n E_\nu^\pi\left[\omega_3^2(x(t))\right]\mathrm{d}t\\ &\le \frac{1}{n}\int_0^n\sum_{i\in X}\left[\mathrm{e}^{-\kappa_1 t}\omega_3^2(i)+\frac{\gamma_1}{\kappa_1}(1-\mathrm{e}^{-\kappa_1 t})\right]\nu(i)\mathrm{d}t\\ &\le \nu(\omega_3^2)+\frac{\gamma_1}{\kappa_1}<\infty. \end{aligned}\tag{8.39}$$

On the other hand, from conditions (D1) and (D4), we see that the sets $\{(i,a)\in K : \omega_3^2(i)\le z\omega_3(i)\}$ are compact in K for each $z\ge 1$. Hence, by (8.39) and Corollary A.46 in [14, p.424], we conclude that the sequence $\{\mu_n\}$ is relatively compact in the ω_3-weak topology. Thus, there exist a subsequence $\{\mu_{n_l}\}$ of $\{\mu_n\}$ and a probability measure $\mu\in\mathscr{M}_{\omega_3}(X\times A)$ such that μ_{n_l} converges to μ in the ω_3-weak topology. By condition (D4), we see that K is closed. Then, since μ_{n_l} weakly converges to μ, by Theorem A.38 in [14, p.420], we have

$$0=\liminf_{l\to\infty}\mu_{n_l}(K^c)\ge\mu(K^c)\ge 0,$$

which implies $\mu(K)=1$. Moreover, for each bounded function v on X, a direct calculation together with condition (D5), the Fubini theorem, the Kolmogorov forward equation, and Theorem 2.3 in [20, p.14] gives

$$\begin{aligned}&\sum_{i\in X}\int_A\left[\sum_{j\in X}q(j|i,a)v(j)\right]\mu_{n_l}(i,\mathrm{d}a)\\ &=\frac{1}{n_l}\int_0^{n_l}\sum_{i\in X}\int_A\left[\sum_{j\in X}q(j|i,a)v(j)\right]P_\nu^\pi(x(t)=i,a(t)\in\mathrm{d}a)\mathrm{d}t\\ &=\frac{1}{n_l}\int_0^{n_l}\sum_{k\in X}\sum_{i\in X}\int_A\left[\sum_{j\in X}q(j|i,a)v(j)\right]P_k^\pi(x(t)=i)\pi_t(\mathrm{d}a|i)\nu(k)\mathrm{d}t\\ &=\frac{1}{n_l}\int_0^{n_l}\sum_{k\in X}\sum_{j\in X}v(j)\left[\sum_{i\in X}q(j|i,\pi_t)p_\pi(0,k,t,i)\right]\nu(k)\mathrm{d}t\\ &=\frac{1}{n_l}\int_0^{n_l}\sum_{k\in X}\sum_{j\in X}v(j)\frac{\partial p_\pi(0,k,t,j)}{\partial t}\nu(k)\mathrm{d}t\\ &=\frac{1}{n_l}\sum_{k\in X}\sum_{j\in X}v(j)\left[p_\pi(0,k,n_l,j)-p_\pi(0,k,0,j)\right]\nu(k)\\ &=\frac{1}{n_l}E_\nu^\pi\left[v(x(n_l))\right]-\frac{1}{n_l}\sum_{k\in X}v(k)\nu(k), \end{aligned}\tag{8.40}$$

where $p_\pi(0,i,t,j)$ denotes the minimal transition function with transition rates $q(j|i,\pi_t) := \int_{A(i)} q(j|i,a)\pi_t(\mathrm{d}a|i)$ for all $i,j \in X$ and $t \geq 0$. Letting $l \to \infty$ in (8.40), by conditions (D5), (D7), and (8.39), we have

$$\sum_{i\in X}\int_A \left[\sum_{j\in X} q(j|i,a)v(j)\right]\mu(i,\mathrm{d}a) = 0$$

for each bounded function v on X, which together with Lemma 4.6 in [30] yields $\mu \in \mathscr{N}_1$. Hence, there exists $\varphi_0 \in \Phi$ such that $\mu = \widehat{\mu}_{\varphi_0}$. Furthermore, by condition (D3), (8.38), and (8.39), we have

$$\overline{V}(c,\pi) \geq \limsup_{n\to\infty} \sum_{i\in X}\int_A c(i,a)\mu_n(i,\mathrm{d}a) \quad \text{and} \quad \overline{V}(r,\pi) \leq \liminf_{n\to\infty} \sum_{i\in X}\int_A r(i,a)\mu_n(i,\mathrm{d}a),$$

which together with conditions (D3) and (D7) yield

$$\rho \geq \overline{V}(c,\pi) \geq \sum_{i\in X}\int_A c(i,a)\widehat{\mu}_{\varphi_0}(i,\mathrm{d}a) = \overline{V}(c,\varphi_0),$$

and

$$\overline{V}(r,\pi) \leq \sum_{i\in X}\int_A r(i,a)\widehat{\mu}_{\varphi_0}(i,\mathrm{d}a) = \overline{V}(r,\varphi_0).$$

This completes the proof of the lemma. □

By Lemma 8.4.9 we see that the constrained optimization problem (8.35) is equivalent to the following form:

$$\sup_{\varphi\in\Phi} \overline{V}(r,\varphi) \quad \text{subject to} \quad \overline{V}(c,\varphi) \leq \rho. \tag{8.41}$$

Under conditions (D1)–(D8), from (8.41) we define a real-valued function G on $C \times D := C_{\omega_3}(K) \times \Phi$ as follows:

$$G(c,\varphi) := \overline{V}(c,\varphi) \quad \text{for} \quad (c,\varphi) \in C \times D = C_{\omega_3}(K) \times \Phi. \tag{8.42}$$

Then, by (8.36) and (8.37), we see that the function G defined in (8.42) satisfies (8.1). Moreover, let $D' := F$.

Now we state our main result of Sect. 8.4.3 on the existence of constrained-optimal policies for (8.35).

Proposition 8.4.2. *Suppose that conditions (D1)–(D9) hold. Then there exists a constrained-optimal policy for (8.35), which may be a stationary policy or a randomized stationary policy that randomizes between two stationary policies differing in at most one state.*

Proof. It follows from Lemma 7.2, Theorem 7.8, and Lemma 12.5 in [20] that Assumptions 8.2.1 and 8.3.2 hold. Obviously, condition (D9) implies Assumption 8.2.2. Finally, from Lemma 12.6 and the proof of Theorem 12.4 in [20], we see that Assumption 8.3.1 holds. Hence, by Theorem 8.3.2, we complete the proof. □

Remark 8.4.10. Proposition 8.4.2 shows that the nonnegativity assumption on the costs as in [20, 35] is not required.

8.4.4 Continuous-Time Constrained MDPs with Discounted Criteria

In this subsection, we consider the following discounted criteria $J(r,\pi)$ and $J(c,\pi)$ in (8.43) below for the model (8.34), in lieu of the average criteria above. Removing the nonnegativity assumption on the cost function as in [18, 20], we will prove that the corresponding results in [18, 20] still hold using Theorem 8.3.2 above.

With the same components as in the model (8.34), we consider the following drift condition from [18, 20]:

(E1) There exists a function $\omega_4 \geq 1$ on X and constants $\gamma_2 \geq 0$, $\kappa_2 \neq 0$, and $\overline{L} > 0$ such that

$$q^*(i) \leq \overline{L}\omega_4(i) \text{ and } \sum_{j\in X} \omega_4(j)q(j|i,a) \leq \kappa_2\omega_4(i) + \gamma_2 \text{ for all } (i,a) \in K.$$

Fix a discount factor $\alpha > 0$ and an initial distribution ν on X. For each $\pi \in \Pi_m$, we define the discounted criteria, $J(r,\pi)$ and $J(c,\pi)$, as follows:

$$J(r,\pi) := E_\nu^\pi\left[\int_0^\infty \mathrm{e}^{-\alpha t} r(x(t),a(t))\mathrm{d}t\right],\ J(c,\pi) := E_\nu^\pi\left[\int_0^\infty \mathrm{e}^{-\alpha t} c(x(t),a(t))\mathrm{d}t\right]. \tag{8.43}$$

Then, the constrained optimization problem for the discounted criteria is as follows:

$$\sup_{\pi\in\Pi_m} J(r,\pi) \text{ subject to } J(c,\pi) \leq \rho. \tag{8.44}$$

To ensure the existence of a constrained-optimal policy π^* for (8.44) (i.e., $J(r,\pi^*) \geq J(r,\pi)$ for all π such that $J(c,\pi) \leq \rho$), we consider the following conditions from [18, 20]:

(E2) There exists a constant $L_4 > 0$ such that

$$|r(i,a)| \leq L_4\omega_4(i) \text{ and } |c(i,a)| \leq L_4\omega_4(i) \text{ for all } (i,a) \in K.$$

(E3) The positive discount factor α verifies that $\alpha > \kappa_2$, with κ_2 as in (E1).
(E4) $v(\omega_4) := \sum_{i \in X} \omega_4(i) v(i) < \infty$.
(E5) For each $i \in X$, $A(i)$ is compact.
(E6) For each $i, j \in X$, the functions $r(i,a)$, $c(i,a)$, $q(j|i,a)$ and $\sum_{k \in X} \omega_4(k) q(k|i,a)$ are continuous in $a \in A(i)$.
(E7) There exist a nonnegative function ω' on X and constants $\gamma_3 \geq 0$, $\kappa_3 > 0$, and $L' > 0$ such that

$$q^*(i)\omega_4(i) \leq L'\omega'(i) \quad \text{and} \quad \sum_{j \in X} \omega'(j) q(j|i,a) \leq \kappa_3 \omega'(i) + \gamma_3 \quad \text{for all } (i,a) \in K.$$

(E8) There exists $\widehat{\pi} \in \Pi_m$ such that $J(c, \widehat{\pi}) < \rho$.

Note that the set Π_m is convex. That is, if π^1 and π^2 are in Π_m, and for any $p \in [0,1]$, $i \in X$, and $t \in [0, \infty)$, $\pi_t^p(\cdot|i) := p\pi_t^1(\cdot|i) + (1-p)\pi_t^2(\cdot|i)$, then $\pi^p \in \Pi$.

Under conditions (E1)–(E6), we define a real-valued function G on $C \times D := C_{\omega_4}(K) \times \Pi_m$ as follows:

$$G(c, \pi) := J(c, \pi), \quad \text{for } (c, \pi) \in C \times D = C_{\omega_4}(K) \times \Pi_m. \tag{8.45}$$

Obviously, the function G defined in (8.45) satisfies (8.1). Moreover, let $D' := F$.

Now we state our main result of Sect. 8.4.4 on the existence of constrained-optimal policies for (8.44).

Proposition 8.4.3. *Suppose that conditions (E1)–(E8) hold. Then there exists a constrained-optimal policy for (8.44), which may be a stationary policy or a randomized stationary policy that randomizes between two stationary policies differing in at most one state.*

Proof. It follows from Theorems 6.5 and 6.10 and Lemma 11.6 in [20] that Assumptions 8.2.1 and 8.3.2 hold. Obviously, condition (E8) implies Assumption 8.2.2. Finally, from Lemma 11.7 and the proof of Theorem 11.4 in [20], we see that Assumption 8.3.1 holds. Hence, by Theorem 8.3.2, we complete the proof. □

Remark 8.4.11. Proposition 8.4.3 shows that the nonnegativity assumption on the costs as in [18, 20] is not required.

References

1. Aliprantis, C., Border, K. (2007). *Infinite Dimensional Analysis*. Springer-Verlag, New York.
2. Altman, E. (1999). *Constrained Markov Decision Processes*. Chapman and Hall/CRC Press, London.
3. Beutler, F.J., Ross, K.W. (1985). Optimal policies for controlled Markov chains with a constraint. *J. Math. Anal. Appl.* **112**, 236–252.

4. Borkar, V., Budhiraja, A. (2004). Ergodic control for constrained diffusions: characterization using HJB equations. *SIAM J. Control Optim.* **43**, 1467–1492.
5. Borkar, V.S., Ghosh, M.K. (1990). Controlled diffusions with constraints. *J. Math. Anal. Appl.* **152**, 88–108.
6. Borkar, V.S. (1993). Controlled diffusions with constraints II. *J. Math. Anal. Appl.* **176**, 310–321.
7. Borkar, V.S. (2005). Controlled diffusion processes. *Probab. Surv.* **2**, 213–244.
8. Budhiraja, A. (2003). An ergodic control problem for constrained diffusion processes: existence of optimal Markov control. *SIAM J. Control Optim.* **42**, 532–558.
9. Budhiraja, A., Ross, K. (2006). Existence of optimal controls for singular control problems with state constraints. *Ann. Appl. Probab.* **16**, 2235–2255.
10. Feinberg, E.A., Shwartz, A. (1995). Constrained Markov decision models with weighted discounted rewards. *Math. Oper. Res.* **20**, 302–320.
11. Feinberg, E.A., Shwartz, A. (1996). Constrained discounted dynamic programming. *Math. Oper. Res.* **21**, 922–945.
12. Feinberg, E.A., Shwartz, A. (1999). Constrained dynamic programming with two discount factors: applications and an algorithm. *IEEE Trans. Autom. Control.* **44**, 628–631.
13. Feinberg, E.A. (2000). Constrained discounted Markov decision processes and Hamiltonian cycles. *Math. Oper. Res.* **25**, 130–140.
14. Föllmer, H., Schied, A. (2004). *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, Berlin.
15. González-Hernández, J., Hernández-Lerma, O. (1999). Envelopes of sets of measures, tightness, and Markov control processes. *Appl. Math. Optim.* **40**, 377–392.
16. González-Hernández, J., Hernández-Lerma, O. (2005). Extreme points of sets of randomized strategies in constrained optimization and control problems. *SIAM J. Optim.* **15**, 1085–1104.
17. Guo, X.P. (2000). Constrained denumerable state non-stationary MDPs with expected total reward criterion. *Acta Math. Appl. Sin. (English Ser.)* **16**, 205–212.
18. Guo, X.P., Hernández-Lerma, O. (2003). Constrained continuous-time Markov controlled processes with discounted criteria. *Stochastic Anal. Appl.* **21**, 379–399.
19. Guo, X.P. (2007). Constrained optimization for average cost continuous-time Markov decision processes. *IEEE Trans. Autom. Control.* **52**, 1139–1143.
20. Guo, X.P., Hernández-Lerma, O. (2009). *Continuous-time Markov Decision Processes: Theory and Applications.* Springer-Verlag, Berlin Heidelberg.
21. Guo, X.P., Song, X.Y. (2011). Discounted continuous-time constrained Markov decision processes in Polish spaces. *Ann. Appl. Probab.* **21**, 2016–2049.
22. Hernández-Lerma, O., Lasserre, J.B. (1996). *Discrete-Time Markov Control Processes: basic optimality criteria.* Springer-Verlag, New York.
23. Hernández-Lerma, O., Lasserre, J.B. (1999). *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, New York.
24. Hernández-Lerma, O., González-Hernández, J. (2000). Constrained Markov control processes in Borel spaces: the discounted case. *Math. Methods Oper. Res.* **52**, 271–285.
25. Huang, Y., Kurano, M. (1997). The LP approach in average reward MDPs with multiple cost constraints: the countable state case. *J. Inform. Optim. Sci.* **18**, 33–47.
26. González-Hernández, J., López-Martínez, R.R., Pérez-Hernández, J.R. (2007). Markov control processes with randomized discounted cost. *Math. Methods. Oper. Res.* **65**, 27–44.
27. López-Martínez, R.R., Hernández-Lerma, O. (2003). The Lagrange approach to constrained Markov processes: a survey and extension of results. *Morfismos.* **7**, 1–26.
28. Mey, S.P., Tweedie, R.L. (1993). Stability of Markov processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Prob.* **25**, 518–548.
29. Piunovskiy, A.B. (1997). *Optimal Control of Random Sequences in Problems with Constraits.* Kluwer, Dordrecht.
30. Prieto-Rumeau, T., Hernández-Lerma, O. (2006). Ergodic control of continuous-time Markov chains with pathwise constraints. *SIAM J. Control Optim.* **45**, 51–73.

31. Puterman, M.L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley, New York.
32. Sennott, L.I. (1991). Constrained discounted Markov chains. *Probab. Engrg. Inform. Sci.* **5**, 463–475.
33. Sennott, L.I. (1993). Constrained average cost Markov chains. *Probab. Engrg. Inform. Sci.* **7**, 69–83.
34. Wei, Q.D., Guo, X.P. (2011). Markov decision processes with state-dependent discount factors and unbounded rewards/costs. *Oper. Res. Lett.* **39**, 369–374.
35. Zhang, L.L., Guo, X.P. (2008). Constrained continuous-time Markov decision processes with average criteria. *Math. Methods Oper. Res.* **67**, 323–340.
36. Zhang, Y. (2011). Convex analytic approach to constrained discounted Markov decision processes with non-constant discount factors. *Top.* doi: 10.1007/s11750-011-0186-8.