The Annals of Applied Probability
2011, Vol. 21, No. 5, 2016–2049
DOI: 10.1214/10-AAP749

DISCOUNTED CONTINUOUS-TIME CONSTRAINED MARKOV DECISION PROCESSES IN POLISH SPACES[1]

BY XIANPING GUO AND XINYUAN SONG

Zhongshan University and The Chinese University of Hong Kong

This paper is devoted to studying constrained continuous-time Markov decision processes (MDPs) in the class of randomized policies depending on *state histories*. The transition rates may be *unbounded*, the reward and costs are admitted to be *unbounded from above and from below*, and the state and action spaces are Polish spaces. The optimality criterion to be maximized is the expected discounted rewards, and the constraints can be imposed on the expected discounted costs. First, we give conditions for the nonexplosion of underlying processes and the finiteness of the expected discounted rewards/costs. Second, using a technique of occupation measures, we prove that the constrained optimality of continuous-time MDPs can be transformed to an *equivalent* (optimality) problem over a class of probability measures. Based on the equivalent problem and a so-called $\bar{w}$*-weak convergence* of probability measures developed in this paper, we show the existence of a constrained optimal policy. Third, by providing a linear programming formulation of the equivalent problem, we show the solvability of constrained optimal policies. Finally, we use two *computable* examples to illustrate our main results.

1. Introduction. *Constrained* Markov decision processes (MDPs) form an important class of stochastic control problems and have been widely studied. Existing works on constrained MDPs can be roughly classified into four groups: (i) constrained discrete-time MDPs with denumerable states [1, 2, 6–10, 23, 25, 37, 38, 41] and their extensive references, (ii) constrained discrete-time MDPs with a Polish state space [19, 20, 29, 33] and their bibliographies, (iii) constrained continuous-time MDPs with denumerable states [13, 15, 34, 36, 42], and (iv) constrained continuous-time MDPs with a Polish state space [11]. A review of these references shows that most of the related literature is concentrated with the first three groups. To the best of our knowledge, the fourth group is addressed only in [11] for the average criteria. Concerning group (i), the existence and algorithms of constrained optimal policies are given in [6–10] for *variant* discounted criteria when states and actions are finite, in [1, 25, 37] for the discounted criteria and denumerable states, and in [1, 2, 23, 37, 38] for the average criteria and denumerable

Received June 2010; revised October 2010.

[1]Supported by NSFC, GDUPS(2010) and GRF(450508) from the Research Grant Council of HKSAR.

MSC2010 subject classifications. 90C40, 60J27.

Key words and phrases. Continuous-time Markov decision process, unbounded transition rates, occupation measure, linear programming formulation, constrained optimal policy.

states. Also, the existence of constrained optimal policies and linear programming formulation for group (ii) are given in [19, 33] for the discounted criteria and in [20, 29, 33] for the average criteria. Although group (iii) has been studied in [13, 15, 34, 36, 42], the references [13, 15, 34, 36, 42] deal with the case of a single constraint, the transition rates in [34] are assumed to be bounded, and the assumption of denumerable states in these references cannot be dropped. On the other hand, as mentioned above, constrained MDPs in Polish spaces are also studied in [19, 20, 29, 33] for the discrete-time case and in [11] for the continuous-time case. However, the reward and cost functions in [29] are assumed to be all *bounded*, and all cost functions in [11, 19, 20, 33] are assumed to be essentially *nonnegative*. Further, such nonnegativeness assumption cannot be removed because it is required for the use of the standard weak convergence of probability measures. This in turn implies that the constrained optimality problem of minimizing nonnegative costs in [11, 19, 20] with constraints imposed on other nonnegative costs cannot be transformed to an equivalent optimality problem of maximizing *bounded* rewards as in [29] with constraints imposed on bounded costs. Hence, the constrained discrete and continuous time MDPs with Polish spaces, in which rewards (to be maximized) and costs (with constraints) may be unbounded from above and from below, have not been studied.

On the other hand, as is known, continuous-time MDPs in Polish spaces have been studied in [11, 12, 16, 27, 34]. However, the treatments in [12, 16, 27] are on the unconstrained case, whereas the results in [11] for the constrained case cannot be applied to the case in which the criterion to be maximized is *unbounded* rewards. This is because the cost to be minimized in [11] is required to be *nonnegative*. Moreover, the study in [11, 12, 16] with unbounded transition rates is limited to the class of *Markov* policies, and yet the case of randomized policies depending on state histories in [27, 34] is for *bounded* transition rates. Hence, as noted in [15, 17, 40], the study on unconstrained continuous-time MDPs with unbounded transition rates and history-dependent policies is an unsolved problem.

Constrained continuous-time MDPs with unbounded transition rates and policies depending on state histories have *not* been studied yet, and they will be considered in this paper. More precisely, we will deal with constrained continuous-time MDPs, which have the following features: (1) the transition rates may be *unbounded*; (2) the reward and costs are admitted to be unbounded *from above and from below*; (3) the state and action spaces are Polish spaces; (4) admissible policies can be randomized and *depend on state histories*; and (5) the optimality criterion is to *maximize* expected discounted rewards, and several constraints are imposed on expected discounted costs.

First, we give the conditions under which we ensure the nonexplosion of underlying processes induced from *unbounded* transition rates and randomized policies *depending on state histories* (see Theorem 3.1 below). This result is a natural extension of the corresponding regularity of a jump Markov process in [5, 12, 15,

16, 31] to a so-called "non-Markov" case and also a generalization of the regularity in [18, 26–28, 30, 34, 37, 39, 40] for *bounded* transition rates. Inspired by the condition for the nonexplosion, we obtain a condition (see Theorem 3.3 below) for the finiteness of the expected discount rewards/costs of each policy when rewards/costs are *unbounded*.

Second, as in [1, 2, 19–21, 29, 33, 35] for constrained MDPs, by introducing an occupation measure, we prove that the constrained optimality problem in continuous-time MDPs [see (2.12) below] can be transformed into an *equivalent* optimality problem [see (3.3) below] over a class of some probability measures. The standard weak convergence technique used in [11, 19, 20, 22, 27, 29] for nonnegative costs does not apply directly to the case wherein rewards/costs are unbounded from above and from below. Therefore, to solve the equivalent optimality problem in which rewards/costs may be unbounded from above and from below, we introduce (Definition 3.7 below) a so-called $\bar{w}$-*weak convergence* of probability measures. This $\bar{w}$-weak convergence is an extension of the standard weak convergence of probability measures. Using the properties of the $\bar{w}$-weak convergence and occupation measures developed here (see Theorem 3.5 and Lemmas 3.8 and 3.9 below), we prove the existence of a constrained optimal policy under mild reasonable conditions (see Theorem 3.11 below). These conditions are slightly different from the usual continuity-compactness ones in [12–15] for continuous-time MDPs and in [1, 2, 19, 20, 22, 29] for the discrete-time MDPs, and thus they are weaker than those in the literature [12–15, 37]; see Remarks 3.10 and 3.12 for details.

Third, for the solvability of constrained optimal policies, we further transform the equivalent optimality problem to a linear programming (LP) problem [see (3.9) below] by using the properties of occupation measures again. Then we present the relationship between a constrained optimal policy and an optimal solution to the LP (see Theorem 3.13 below), and characterize a stationary policy (see Theorem 3.15 below). This relationship and characterization of a stationary policy are used to obtain the solvability and structure of a constrained optimal policy (see Corollary 3.14 and Theorem 3.16 below).

Finally, to illustrate our main results, we present two *computable* examples in which our conditions are satisfied, whereas some of those in [11, 19, 20, 22, 27, 29] fail to hold (see Remark 4.7 below). In particular, our approach is also *suitable* to the case of discrete-time MDPs with rewards/costs being unbounded from above and from below, and similar results for the discrete-time case can also be obtained; see Remark 3.17 for details. However, our model *cannot* be transformed to an equivalent one of discrete-time MDPs using the uniformization technique because the transition rates in our model may be *unbounded*.

The rest of this paper is organized as follows. In Section 2, the model and the constrained optimality problem that we are concerned with are introduced. The main results of this paper are stated in Section 3, and illustrated with computable examples in Section 4. The proofs of the main results are presented in Section 5.

2. The model for constrained continuous-time MDPs.

Notation. If X is a Polish space (i.e., a complete and separable metric space) and $\bar{w} \geq 1$ is a real-valued measurable function on X, we denote by $\mathcal{B}(\mathrm{X})$ the Borel σ-algebra on X, by D^c the complement of a set $D \subseteq X$ (with respect to X), by $\|u\|_{\bar{w}}$ the $\bar{w}$-weighted norm of a real-valued measurable function u on X [i.e., $\|u\|_{\bar{w}} := \sup_{x \in X} |u(x)|/\bar{w}(x)$], by $C_b(X)$ the set of all bounded continuous functions on X, and by $\mathcal{P}(X)$ the set of all probability measures on $\mathcal{B}(\mathrm{X})$. Let

$$B_{\bar{w}}(X) := \{u | \|u\|_{\bar{w}} < \infty\}$$

be the Banach space.

We now introduce the model of constrained continuous-time MDPs,

$$\{S, (A(x) \subseteq A, x \in S), q(\cdot|x,a), r(x,a), (c_n(x,a), d_n, 1 \leq n \leq N)\}, \tag{2.1}$$

where S is a *state space*, A is an *action space*, and $A(x)$ is a Borel *set of admissible actions* at state $x \in S$. We suppose that S and A are Polish spaces, and the following set:

$$K := \{(x,a) | x \in S, a \in A(x)\} \tag{2.2}$$

is a Borel subset of $S \times A$.

The function $q(\cdot|x,a)$ in (2.1) refers to *transition rates*, that is, it satisfies the following:

(T_1) For each fixed $(x,a) \in K$, $q(\cdot|x,a)$ is a *signed* measure on $\mathcal{B}(S)$, whereas for each fixed $D \in \mathcal{B}(S)$, $q(D|\cdot)$ is a real-valued Borel-measurable function on K;

(T_2) $0 \leq q(D|x,a) < \infty$ for all $(x,a) \in K$ and $x \notin D \in \mathcal{B}(S)$; and

(T_3) $q(S|x,a) = 0$ for all $(x,a) \in K$. [Hence, $q(\{x\}|x,a)$ is finite for all $(x,a) \in K$.]

The model is also assumed to be *stable*, which means

$$q^*(x) := \sup_{a \in A(x)} |q(\{x\}|x,a)| < \infty \qquad \forall x \in S. \tag{2.3}$$

Finally, the function $r(x,a)$ on K denotes the reward, whereas the functions $c_n(x,a)$ on K and the real numbers d_n denote the costs and constraints, respectively. We assume that $r(x,a)$ and $c_n(x,a)$ are real-valued measurable on K. [$r(x,a)$ *is allowed to take positive and negative values, so it can be interpreted as a cost rather than a "reward" only.*]

To complete the specification of the constrained optimality problem, we of course need an optimality criterion. This requires the definition of a class of policies admissible to a controller. To do so, we introduce some notation as in [24, 27, 28].

Let $S_\infty := S \cup \{x_\infty\}$ with x_∞ being an isolated point, $\Omega^0 := (S \times \mathbb{R}_+)^\infty$ with $\mathbb{R}_+ := (0, \infty)$ and $\Omega := \Omega^0 \cup \{(x_0, \theta_1, x_1, \ldots, \theta_{k-1}, x_{k-1}, \infty, x_\infty, \ldots) | \theta_l \in$

$\mathbb{R}_+, x_0, x_l \in S$ for each $1 \le l \le k-1$ and $k \ge 2\}$. By the corresponding modification of the σ-algebra over Ω^0, we can obtain the basic measurable space $(\Omega, \mathcal{F})$. Then we define maps T_k, X_k, Θ_k $(k = 0, 1, \ldots)$ and ξ_t $(t \ge 0)$ on $(\Omega, \mathcal{F})$ as follows: for each $e := (x_0, \theta_1, x_1, \ldots, \theta_k, x_k, \ldots) \in \Omega$, let

$$
\begin{aligned}
T_k(e) &:= \theta_1 + \cdots + \theta_k \qquad (\text{for } k \ge 1),\\
T_\infty(e) &:= \lim_{k\to\infty} T_k(e) \qquad \text{with } T_0(e) := 0;\\
X_{k-1}(e) &:= x_{k-1}, \qquad \Theta_k(e) := \theta_k \qquad \text{for } k \ge 1;
\end{aligned} \tag{2.4}
$$

$$
\xi_t(e) := \sum_{k\ge 0} x_k I_{\{T_k \le t < T_{k+1}\}}(e) + x_\infty I_{\{T_\infty \le t\}}(e), \tag{2.5}
$$

where I_D stands for the indicator function of a set D. Let $h_k(e) = (x_0, \theta_1, x_1, \ldots, \theta_k, x_k)$, and call $h_k(e)$ a k-component *state history*. Obviously, these maps are measurable on $\mathcal{F}$. In what follows, the argument $e = (x_0, \theta_1, x_1, \ldots, \theta_k, x_k, \ldots)$ is often omitted.

Components Θ_k play the role of inter-jump intervals or sojourn times, T_k are the jump epoches, and X_k denotes the state of the process $\{\xi_t, t \ge 0\}$ on $[T_k, T_{k+1})$. We do not intend to consider the process after moment T_∞, so we view it to be absorbed in state x_∞. Hence, we write $q(\cdot|x_\infty, a_\infty) \equiv 0$, where a_∞ is an isolated point, and let $A(x_\infty) := \{a_\infty\}$, $A_\infty := A \cup \{a_\infty\}$.

Let $\mathbb{R}^0_+ := [0, \infty)$, and introduce the integer-valued random measure μ^* on $\mathbb{R}^0_+ \times S$ by

$$
\mu^*(dt, dx) = \sum_{k\ge 0} I_{\{T_k < \infty\}} \delta_{(T_k, X_k)}(dt, dx), \tag{2.6}
$$

where $\delta_y(\cdot)$ is the Dirac measure concentrated at any point y. Then we take the right-continuous family of σ-algebras $\{\mathcal{F}_t\}_{t\ge 0}$ with $\mathcal{F}_t := \sigma\{\mu^*([0, s] \times D), s \in [0, t], D \in \mathcal{B}(S)\}$, and let

$$
\mathcal{P} := \sigma\big(B \times \{0\}, C \times (s, \infty) | B \in \mathcal{F}_0, C \in \mathcal{F}_{s-}, s > 0\big),
$$

where $\mathcal{F}_{s-} := \bigvee_{t<s} \mathcal{F}_t$. Then, as in [24, 27, 28], a real-valued function on $\Omega \times \mathbb{R}^0_+$ is called *predictable* if it is measurable with respect to $\mathcal{P}$.

We next introduce the definition of a policy, which is the same as in [27] and a generalization of the corresponding one in [28, 34, 35] for denumerable states.

DEFINITION 2.1. A transition probability π from $(\Omega \times \mathbb{R}^0_+, \mathcal{P})$ onto $(A_\infty, \mathcal{B}(A_\infty))$ such that $\pi(A(\xi_{t-}(e))|e, t) \equiv 1$ is called a policy, which can be randomized and depend on state histories. A policy is called *randomized stationary* if there exists a transition probability ϕ from $(S, \mathcal{B}(S))$ onto $(A, \mathcal{B}(A))$ such that $\phi(A(x)|x) \equiv 1$ and $\pi(da|e, t) = I_{\{t<T_\infty\}}(e)\phi(da|\xi_{t-}(e)) + I_{\{t\ge T_\infty\}}(e)\delta_{a_\infty}(da)$. We will write such a randomized stationary policy as ϕ. A randomized stationary

policy ϕ is called (deterministic) *stationary* if there exists a measurable function f from $(S, \mathcal{B}(S))$ onto $(A, \mathcal{B}(A))$ such that $\phi(\{f(x)\}|x) \equiv 1$. Such a stationary policy will be written as f.

We denote by Π, Π_s and F the classes of all policies, randomized stationary policies and stationary policies, respectively. Equivalently, Π_s is the set of all stochastic kernels ϕ on A given S such that $\phi(A(x)|x) = 1$ for all $x \in S$, and F is the set of all measurable functions f from S to A such that $f(x) \in A(x)$ for all $x \in S$. Obviously, $F \subset \Pi_s \subset \Pi$.

REMARK 2.2. The requirement of predictability of a policy implies that at time $t \geq 0$ each policy depends on only the past jump moments $T_0, T_1, \ldots, T_m \leq t$ and the corresponding states $x_0, \ldots, x_m \in S$. This means that a policy may depend on state histories. However, the class Π is not the complete collection of all history-dependent policies. This is because each state history $h_k = (x_0, \theta_1, x_1, \ldots, \theta_k, x_k)$ does not include past actions a_m $(0 \leq m \leq k)$. To overcome the shortcoming of the definition of a state history, a possible and natural way is to replace h_k with a new history $(x_0, a_0, \theta_1, \ldots, x_{k-1}, a_{k-1}, \theta_k, x_k)$ including past actions. If we do so, some results in [24, 28] such as the structure of the probability measure P_γ^π in (2.9) and the predictable properties of the randomized measure ν^π in (2.7) and functions $m(D|e,t)$ in (2.8), which are required in following arguments, need to be checked one by one. Since these desired results for the case of new histories have not been proven, we still use the definition of a policy in Definition 2.1, which is the same as in [27, 28, 34, 35], and which is also a generalization of the corresponding one in [5, 11, 12, 15, 17] for a Markov policy.

For each $\pi \in \Pi$, by Definition 2.1 we see that the random measure on $\mathbb{R}_+^0 \times S$ given by

$$\nu^\pi(e, dt, D) := \left[\int_A \pi(da|e,t) q(D|\xi_{t-}(e), a) I_{\{\xi_{t-} \notin D\}}(e)\right] dt \quad \text{for } D \in \mathcal{B}(S) \tag{2.7}$$

is predictable, and $\nu^\pi(\{t\} \times S) = \nu^\pi([T_\infty, \infty) \times S) \equiv 0$ for all $t \geq 0$. Thus, for any initial distribution $\gamma \in \mathcal{P}(S)$, Theorem 4.27 in [28] (or Theorem 3.6 in [24]) ensures the existence of a unique probability measure P_γ^π on $(\Omega, \mathcal{F})$ such that $P_\gamma^\pi\{x_0 \in dx\} = \gamma(dx)$, and ν^π is a dual predictable projection of the measure μ^* in (2.6). The expectation operator with respect to P_γ^π is denoted by E_γ^π. In particular, E_γ^π and P_γ^π will be written as E_x^π and P_x^π, respectively, when γ is the Dirac measure located at point $x \in S$.

For any fixed $\pi \in \Pi$ and $\gamma \in \mathcal{P}(S)$, let us recall how the measure P_γ^π is constructed. First, by Definition 2.1 we see that, for each fixed $D \in \mathcal{B}(S)$, the following

function on $\Omega \times \mathbb{R}_+^0$:

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

is predictable, and thus (by Lemma 3.3 in [24]) has the following representation:

$$\begin{aligned} m(D|e,t) =: {} & I_{\{0\}}(t) m_0(D|x_0,0) \\ & + \sum_{k=0}^{\infty} I_{\{T_k < t \le T_{k+1}\}}(e) m_k\big(D|h_k(e), t - T_k\big), \end{aligned} \tag{2.8}$$

where $m_k(\cdot|h_k(e),\tilde{t}\,)$ (depending on π) is a measure on $\mathcal{B}(S)$ [for any fixed $h_k(e)$ and $\tilde{t}$], $m_k(D|h_k(e),\tilde{t}\,)$ is measurable in $(e,\tilde{t}\,)$ [for any fixed $D \in \mathcal{B}(S)$] and $m_k(\{x_k\}|h_k(e),\tilde{t}\,) = 0$ for all $x_k \in S$ and $k \ge 0$. Let $\hat{H}_0 \triangleq S$, $\hat{H}_k \triangleq S \times (\mathbb{R}_+ \times S_\infty)^k$ for $k \ge 1$. Noting that a measure γ on $\mathcal{B}(\hat{H}_0)$ is given, we suppose that the measure P_γ^π on $\mathcal{B}(\hat{H}_k)$ has been constructed, then P_γ^π on $\mathcal{B}(\hat{H}_{k+1})$ is determined as follows:

$$\begin{aligned} & P_\gamma^\pi\big(\Gamma \times (d\tilde{t}, dx)\big) \\ & \quad := \int_\Gamma P_\gamma^\pi(dh_k) I_{\{\theta_{k+1} < \infty\}} m_k(dx|h_k,\tilde{t}\,) e^{-\int_0^{\tilde{t}} m_k(S|h_k,v)\,dv}\, d\tilde{t}; \\ & P_\gamma^\pi\big(\Gamma \times (\infty, x_\infty)\big) \\ & \quad := \int_\Gamma P_\gamma^\pi(dh_k)\{I_{\{\theta_{k+1}=\infty\}} + I_{\{\theta_{k+1}<\infty\}} e^{-\int_0^\infty m_k(S|h_k,v)\,dv}\}, \end{aligned} \tag{2.9}$$

where $\Gamma \in \mathcal{B}(\hat{H}_k)$. According to the Ionescu Tulcea theorem in [4], there exists a unique probability measure P_γ^π on $(\Omega, \mathcal{F})$, which has projections onto the spaces of k-component state histories satisfying relations (2.9).

For any given $\gamma \in \mathcal{P}(S)$ and $\pi \in \Pi$, using (2.8) and (2.9), we now give a somewhat informal description of how the process $\{\xi_t, t \ge 0\}$ evolves. Suppose that the process is at state x_k at time $t \in [T_k, T_{k+1})$ $(k \ge 0)$. Then, a transition from x_k to a set D of states occurs with probability $m_k(D|h_k, t - T_k)$, or the process remains at x_k with probability $1 - m_k(S|h_k, t - T_k)\,dt + o(dt)$. In the former case, the sojourn time Θ_{k+1} of $\{\xi_t, t \ge 0\}$ at x_k has a distribution with a so-called "density function" $e^{-\int_0^t m_k(S|h_k,v)\,dv}$.

As mentioned above, we do not intend to consider the process after moment T_∞. Thus, we need to give conditions ensuring the nonexplosion of $\{\xi_t, t \ge 0\}$ [i.e., $P_x^\pi(\xi_t \in S) \equiv 1$]. To do so, we consider the following condition.

ASSUMPTION A. There exist a continuous function $w \ge 1$ on S and constants $\rho, b \ge 0$ and a sequence of nondecreasing subsets $\{S_k\}$ of S, such that:

(1) $\int_S w(y) q(dy|x,a) \le \rho w(x) + b$ for all $(x,a) \in K$;

(2) $\inf_{x\notin S_k} w(x) \uparrow +\infty$ as $k \to \infty$, with $\inf \varnothing := \infty$;
(3) $S_k \uparrow S$ and $\sup_{a\in A(x), x\in S_k} |q(\{x\}|x,a)| < \infty$ for all $k \geq 1$.

REMARK 2.3. We call Assumption A a nonexplosion condition for $\{\xi_t, t \geq 0\}$. Obviously, Assumption A trivially holds when the transition rates are *bounded*; see [18, 26, 27, 30, 34, 37, 39, 40], for instance. Assumption A is similar to those in [5, 11, 12, 15, 17] for Markov policies and unbounded transition rates, and it can be verified with examples in [5, 11, 12, 15, 17] and those below.

Under Assumption A, we see (by Theorem 3.1 below) that $\{\xi_t, t \geq 0\}$ is nonexplosive. Thus, for any fixed discount factor $\alpha > 0$ and an initial distribution $\gamma \in \mathcal{P}(S)$, we define the expected discounted criteria

$$V_\alpha(x,\pi,u) := \int_0^\infty e^{-\alpha t} \int_A E_x^\pi [u(\xi_{t-}, a)\pi(da|e,t)]\, dt,$$
$$V_\alpha(\pi,u) := \int_S V_\alpha(x,\pi,u)\gamma(dx) \tag{2.10}$$

for each $\pi \in \Pi$, $x \in S$ and a measurable function u on K, provided the integrals in (2.10) are well defined.

In particular, let

$$V_r(x,\pi) := V_\alpha(x,\pi,r), \qquad V_r(\pi) := V_\alpha(\pi,r)$$

and

$$V_n(x,\pi) := V_\alpha(x,\pi,c_n), \qquad V_n(\pi) := V_\alpha(\pi,c_n) \qquad \text{for } n = 1,\dots,N.$$

[The finiteness of $V_r(\pi)$ and $V_n(\pi)$ will be ensured in Theorem 3.3 below.]

Let

$$U := \{\pi \,|\, V_n(\pi) \leq d_n, n = 1,\dots,N\} \quad \text{and} \quad V_r(U) := \sup_{\pi\in U} V_r(\pi) \tag{2.11}$$

be the set of constrained policies and the constrained optimal reward value, respectively.

In the following arguments, we assume that the set U is *not* empty, and the discount factor α and the initial distribution γ as well as the numbers d_n are *fixed*.

Then, the constrained optimality problem under consideration is as follows:

$$\text{Maximize } V_r(\pi) \text{ over all } \pi \in U. \tag{2.12}$$

DEFINITION 2.4. A policy $\pi^* \in U$ is said to be constrained optimal if $V_r(\pi^*) = V_r(U)$. When $U = \Pi$, a constrained optimal policy is said to be unconstrained optimal.

The main goal of this paper is to give the conditions for the existence and solvability of a constrained/unconstrained optimal policy.

3. Main results. We state the main results of our work in this section. Their proofs are presented later in Section 5. The main results are given in three subsections.

3.1. *Conditions for nonexplosion and finiteness.* This subsection states the results on the nonexposition of $\{\xi_t, t \geq 0\}$ and finiteness of $V_n(x,\pi)$ and $V_n(\pi)$.

For the nonexposition of $\{\xi_t, t \geq 0\}$, we have the following fact.

THEOREM 3.1. *Suppose that Assumption A holds. Then, for each* $\pi \in \Pi$, $x \in S$ *and* $t \geq 0$:

(a) $P_x^\pi(T_\infty = \infty) = 1$ *and* $P_x^\pi(\xi_t \in S) = 1$.

(b)

$$E_x^\pi[w(\xi_t)] \leq \begin{cases} e^{\rho t} w(x) + \dfrac{b}{\rho}(e^{\rho t} - 1), & \text{if } \rho \neq 0, \\ w(x) + bt, & \text{if } \rho = 0. \end{cases}$$

(c) *The analog of the forward Kolmogorov equation holds*:

$$P_x^\pi(\xi_t \in D) = I_D(x) + E_x^\pi\left[\int_0^t \int_A \pi(da|e,s) q(D|\xi_{s-}(e), a)\, ds\right]$$

for each $D \in \mathcal{B}(S)$ *with* $\sup_{x \in D} q^*(x) < \infty$.

The proof of Theorem 3.1 appears in Section 5.

REMARK 3.2. Theorem 3.1(a) establishes the nonexplosion of $\{\xi_t, t \geq 0\}$ on the probability space $(\Omega, \mathcal{F}, P_x^\pi)$ (for each policy $\pi \in \Pi$ and $x \in S$), and Theorem 3.1 is an extension of the corresponding results in [18, 26, 27, 30, 34, 35, 37, 39, 40] for bounded transition rates and in [5, 11–17, 31] for Markov policies only. The process $\{\xi_t, t \geq 0\}$ may *not* be Markovian because a policy π can depend on state histories.

Inspired by Theorem 3.1, we introduce the following condition.

ASSUMPTION B. Let $c_0(x,a) := -r(x,a)$ for $(x,a) \in K$, and w be as in Assumption A.

(1) There exists a constant $M > 0$ such that, $|c_n(x,a)| \leq M w(x)$ for every $(x,a) \in K$ and $n = 0, 1, \ldots, N$.

(2) The discount factor α satisfies that $\alpha > \rho$, with ρ as in Assumption A.

(3) $\int_S w(x)\gamma(dx) < \infty$.

Then the following fact establishes the finiteness of $V_n(x,\pi)$ and $V_n(\pi)$.

THEOREM 3.3. *Suppose that Assumptions* A *and* B *hold. Then, for each* $\pi \in \Pi$ *and* $x \in S$:

(a) $E_x^\pi[|c_n(\xi_t, a)|\pi(da|e,t)] \le M E_x^\pi[w(\xi_t)]$ *for all* $t \ge 0$ *and* $n = 0, 1, \ldots, N$;

(b) $|V_n(x,\pi)| \le M[\alpha w(x) + b]/[\alpha(\alpha - \rho)]$ *and* $|V_n(\pi)| \le M M_1^*$ *for* $n = 0, 1, \ldots, N$, *where* $V_0(x,\pi) := V_\alpha(x,\pi,c_0)$, $V_0(\pi) := V_\alpha(\pi, c_0)$, $M_1^* := [\alpha \times \int_S w(x)\gamma(dx) + b]/[\alpha(\alpha - \rho)]$.

PROOF. Obviously, this theorem follows from Theorem 3.1(b) and (2.10). □

3.2. *Existence of constrained optimal policies*. This subsection states the main results on the existence of constrained optimal policies.

In order to show the existence of a constrained optimal policy, as in [1, 2, 19–21, 29, 33, 35], we introduce a key concept of an occupation measure of a policy.

DEFINITION 3.4. Fix policies $\pi, \pi_1, \pi_2 \in \Pi$.

(i) The occupation measure of π is a probability measure η^π on $S \times A$ concentred on K, which is defined by

$$\eta^\pi(D \times \Gamma) := \alpha \int_0^\infty e^{-\alpha t} E_\gamma^\pi\big[I_{\{\xi_t \in D\}}(e)\pi(\Gamma|e,t)\big]\,dt \quad \text{with } D \in \mathcal{B}(S), \Gamma \in \mathcal{B}(A). \tag{3.1}$$

(Obviously, η^π concentrates on K and depends on π, α and γ. However, we impress γ and α in the occupation measure for simplicity.)

(ii) Two policies π^1 and π^2 are called equivalent if $\eta^{\pi^1} = \eta^{\pi^2}$.

(iii) We denote by $\hat{\eta}$ the *marginal* (or *projection*) on S of a probability measure η on $S \times A$, and by $\phi^\eta (\in \Pi_s)$ the randomized stationary policy (depending on η), which is determined by the following decomposition of η:

$$\eta(dx, da) = \hat{\eta}(dx)\phi^\eta(da|x). \tag{3.2}$$

Thus, by (3.1) and (2.10), we have $V_\alpha(x,\pi,u) = \frac{1}{\alpha}\int_{S\times A} u(x,a)\eta^\pi(dx,da)$, and we can rewrite (2.12) as an *equivalent* optimality problem:

$$\begin{aligned} &\text{Maximize } \frac{1}{\alpha}\int_K r(x,a)\eta(dx,da) \\ &\text{over } \eta \in \left\{\eta^\pi : \int_K c_n(x,a)\eta^\pi(dx,da) \le \alpha d_n,\ 1 \le n \le N\right\}. \end{aligned} \tag{3.3}$$

To solve problem (3.3), we need to seek a certain compactness structure on the set of all occupation measures. To do so, we require to characterize an occupation measure, and we have the following fact.

THEOREM 3.5. *Under Assumption* A, *the following assertions hold.*

(a) *The occupation measure* η^π *(for each fixed* $\pi \in \Pi$*) satisfies the following equation*:

$$\alpha\hat{\eta}^\pi(D) = \alpha\gamma(D) + \int_{S\times A} q(D|x,a)\eta^\pi(dx,da)$$
$$\forall D \in \mathcal{B}(S) \text{ with } \sup_{x\in D} q^*(x) < \infty.$$

(b) *Conversely, if a probability measure* η *on* $S \times A$ *(concentrated on* K*) satisfies*

$$\alpha\hat{\eta}(D) = \alpha\gamma(D) + \int_{S\times A} q(D|x,a)\eta(dx,da)$$
$$\forall D \in \mathcal{B}(S) \text{ with } \sup_{x\in D} q^*(x) < \infty$$

and $\int_S |q(\{x\}|x,\phi^\eta)|\hat{\eta}(dx) < \infty$*, then* $\eta^{\phi^\eta} = \eta$*, where* ϕ^η *is as in* (3.2).

(c) *If, in addition, Assumptions* B(2) *and* B(3) *are satisfied, and* $q^*(x) \le Lw(x)$ *for all* $x \in S$*, with some constant* $L > 0$*, then* $\phi^{\eta^\phi} = \phi$ *for all* $\phi \in \Pi_s$.

The proof of Theorem 3.5 appears in Section 5.

REMARK 3.6. Theorems 3.5(a) and 3.5(b) are proved in [35] for continuous-time MDPs with uniformly bounded transition rates and in [1, 2, 21] for discrete-time MDPs.

To give a certain convergence of occupation measures, we introduce some notation.

For any real-valued continuous function $\bar{w} \ge 1$ on S, let

$$\mathcal{P}_{\bar{w}}(S\times A) := \left\{\eta \in \mathcal{P}(S\times A)\Big| \int_S \bar{w}(x)\hat{\eta}(dx) < \infty\right\}.$$

Then we define two maps, $T_{\bar{w}}$ and $T'_{\bar{w}}$, as follows:

$$T_{\bar{w}}: \quad \mathcal{P}_{\bar{w}}(S\times A) \longrightarrow \mathcal{P}(S\times A), \qquad \eta \mapsto T_{\bar{w}}(\eta),$$

where $T_{\bar{w}}(\eta)$ is given by

$$T_{\bar{w}}(\eta)(D\times\Gamma) := \frac{\int_D \bar{w}(x)\eta(dx,\Gamma)}{\int_S \bar{w}(x)\hat{\eta}(dx)} \qquad \forall D \in \mathcal{B}(S) \text{ and } \Gamma \in \mathcal{B}(A); \tag{3.4}$$

$$T'_{\bar{w}}: \quad \mathcal{P}(S\times A) \longrightarrow \mathcal{P}_{\bar{w}}(S\times A), \qquad \mu \mapsto T'_{\bar{w}}(\mu),$$

where $T'_{\bar{w}}(\mu)$ is given by

$$T'_{\bar{w}}(\mu)(D\times\Gamma) := \frac{\int_D (1/\bar{w}(x))\mu(dx,\Gamma)}{\int_S (1/\bar{w}(x))\hat{\mu}(dx)} \qquad \forall D \in \mathcal{B}(S) \text{ and } \Gamma \in \mathcal{B}(A). \tag{3.5}$$

[Since $1 \le \bar{w} < \infty$ on S, we have $0 < \int_S \frac{1}{\bar{w}(x)}\mu(dx) \le 1$ for any $\mu \in \mathcal{P}(S)$, and thus the maps $T_{\bar{w}}$ and $T'_{\bar{w}}$ are well defined.]

DEFINITION 3.7. The $\bar{w}$-weak topology on $\mathcal{P}_{\bar{w}}(S \times A)$ is defined by the $\bar{w}$-weak convergence as follows: a sequence $\{\eta_k, k \geq 1\} \subseteq \mathcal{P}_{\bar{w}}(S \times A)$ is called to $\bar{w}$-converge weakly to $\eta \in \mathcal{P}_{\bar{w}}(S \times A)$ (and written as $\eta_k \xrightarrow{\bar{w}} \eta$) if

$$\lim_{k\to\infty} \int_{S\times A} u(x,a)\eta_k(dx,da) = \int_{S\times A} u(x,a)\eta(dx,da)$$

for each continuous function $u(x,a)$ on $S \times A$ such that $|u(x,a)| \leq L_u \bar{w}(x)$ for all $(x,a) \in S \times A$, with some nonnegative constant L_u depending on u.

Obviously, $\eta_k \xrightarrow{\bar{w}} \eta$ implies $\eta_k \xrightarrow{1} \eta$ (the standard weak convergence of probability measures). The following lemma establishes the relationship between $\bar{w}$- and standard weak convergence.

LEMMA 3.8. *For any given real-valued continuous function $\bar{w} \geq 1$ on S, let $\{\eta_k, k = 0, 1, \ldots\} \subset \mathcal{P}_{\bar{w}}(S \times A)$ and $\{\mu_k, k = 0, 1, \ldots\} \subset \mathcal{P}(S \times A)$. Then*:

(a) *$T_{\bar{w}}(\eta) \in \mathcal{P}(S \times A)$ for all $\eta \in \mathcal{P}_{\bar{w}}(S \times A)$ and $T'_{\bar{w}}(\mu) \in \mathcal{P}_{\bar{w}}(S \times A)$ for all $\mu \in \mathcal{P}(S \times A)$;*

(b) *$T'_{\bar{w}}(T_{\bar{w}}(\eta)) = \eta$ for all $\eta \in \mathcal{P}_{\bar{w}}(S \times A)$ and $T_{\bar{w}}(T'_{\bar{w}}(\mu)) = \mu$ for all $\mu \in \mathcal{P}(S \times A)$;*

(c) *$\eta_k \xrightarrow{\bar{w}} \eta_0$ if and only if $T_{\bar{w}}(\eta_k) \xrightarrow{1} T_{\bar{w}}(\eta_0)$;*

(d) *$\mu_k \xrightarrow{1} \mu_0$ if and only if $T'_{\bar{w}}(\mu_k) \xrightarrow{\bar{w}} T'_{\bar{w}}(\mu_0)$.*

The proof of Lemma 3.8 appears in Section 5.

To further analyze the properties of occupation measures, we let

$$\mathcal{M}_o := \left\{ \eta^\pi \,\middle|\, \int_S w(x)\hat{\eta}^\pi(dx) < \infty, \pi \in \Pi \right\} \subseteq \mathcal{P}_w(K) \tag{3.6}$$

(with w as in Assumption A),

$$\mathcal{M}_o^c := \left\{ \eta \in \mathcal{M}_o \,\middle|\, \int_{S\times A} c_n(x,a)\eta(dx,da) \leq \alpha d_n, n = 1, \ldots, N \right\}. \tag{3.7}$$

LEMMA 3.9. *Suppose that Assumptions* A, B(2) *and* B(3) *hold. If, in addition, $q^*(x) \leq Lw(x)$ for all $x \in S$, with some constant $L > 0$, then the following assertions hold*:

(a) *$\mathcal{M}_o$ and $\mathcal{M}_o^c$ are convex.*

(b) *If, in addition, $\int_S g(y)q(dy|x,a)$ is continuous on K for each fixed $g \in C_b(S)$, then $\mathcal{M}_o$ is closed* (*with respect to the w-weak topology*).

The proof of Lemma 3.9 appears in Section 5.

For the solvability of (3.3), by Lemmas 3.8 and 3.9, we introduce the following condition.

ASSUMPTION C. Let w be as in Assumption A.

(1) The functions $c_n(x,a)$ and $\int_S g(y)q(dy|x,a)$ are continuous on K [for each fixed $g \in C_b(S)$ and $0 \le n \le N$].

(2) There exist a measurable function $w' \ge 1$ on S and a nondecreasing sequence of compact sets $K_m \uparrow K$, such that $\lim_{m\to\infty} \inf_{(x,a)\notin K_m} \frac{w(x)}{w'(x)} = \infty$.

(3) There exist a constant $L > 0$ such that $q^*(x) \le Lw(x)$ for all $x \in S$.

REMARK 3.10. Assumption C(2) is slightly different from the compactness condition in [19–22, 29] for discrete-time MDPs and [12, 16] for continuous-time MDPs.

We now state our second main result on the existence of a constrained optimal policy.

THEOREM 3.11. *Suppose that Assumptions A, B and C hold. Then*:

(a) *$\mathcal{M}_o$ and $\mathcal{M}_o^c$ are metrizable and compact (with respect to the w'-weak topology), that is, for any sequence $\{\eta_k, k \ge 1\}$ in $\mathcal{M}_o$ (or $\mathcal{M}_o^c$), there exists a subsequence $\{\eta_{k_m}, m \ge 1\}$ and $\eta_0 \in \mathcal{M}_o$ (or $\mathcal{M}_o^c$) such that such that $\eta_{k_m} \xrightarrow{w'} \eta_0$ as $m \to \infty$.*

(b) *There exists a constrained optimal policy.*

The proof of Theorem 3.11 appears in Section 5.

REMARK 3.12. Theorem 3.11(b) shows the existence of a constrained optimal policy. It should be noted that the conditions for Theorem 3.11(b) are weaker than those in [12–15, 37] for the class of all Markov policies. This is because some assumptions such as the nonnegativity of costs in [13] and the absolute integrability condition in [12, 13, 15] are not required here.

3.3. *Solvability of constrained optimal policies*. This subsection states the results on the solvability of constrained optimal policies.

First, by (3.3) we see that the original constrained optimality problem (2.12) is *equivalent* to the following constrained minimization problem:

$$\text{Minimize } V_0(\pi) \text{ over } \pi \in \{\pi \,|\, V_n(\pi) \le d_n, n = 1, \dots, N\}. \tag{3.8}$$

By (2.10) and (3.1), the problem (3.8) can be rewritten into the following form:

$$\begin{cases} \inf\limits_{\eta\in\{\eta^\pi\,|\,\pi\in\Pi\}} \dfrac{1}{\alpha} \displaystyle\int_{S\times A} c_0(x,a)\eta(dx,da), \\ \text{subject to } \displaystyle\int_{S\times A} c_n(x,a)\eta(dx,da) \le \alpha d_n, \qquad n = 1, \dots, N, \end{cases}$$

which (by Theorem 3.5) is equivalent to the following linear program (LP):

$$\text{LP}: \quad \inf_{\eta} \int_{S\times A} \frac{1}{\alpha} c_0(x,a)\eta(dx,da) \tag{3.9}$$

subject to

$$\begin{cases} \displaystyle\int_{S\times A} c_n(x,a)\eta(dx,da) \le \alpha d_n, \qquad n = 1,\dots,N, \\ \alpha\hat{\eta}(D) = \alpha\gamma(D) + \displaystyle\int_{S\times A} q(D|x,a)\eta(dx,da), \\ \qquad \text{for all } D \in \mathcal{B}(S) \text{ with } \sup_{x\in D} q^*(x) < \infty, \\ \displaystyle\int_S w(x)\hat{\eta}(dx) < \infty, \qquad \eta \in \mathcal{P}(K). \end{cases} \tag{3.9$'$}$$

Obviously, (3.9) is a linear program over the set of probability measures $\eta \in \mathcal{P}(K)$ satisfying (3.9′). We call (3.9) the primal *linear programming formulation* of (2.12).

Thus, we obtain the following result on the solvability of constrained optimal policies.

THEOREM 3.13. *Under Assumptions* A, B *and* C(3), *the following assertions hold.*

(a) *If there exists a feasible solution to LP* (3.9), *then the set* U *of constrained policies is nonempty. Conversely, if* U *is nonempty, then there exists a feasible solution to LP* (3.9).

(b) *If there exists an optimal solution* η^* *to LP* (3.9), *then the randomized stationary policy* ϕ^{η^*} *is constrained optimal. Conversely, if* π^* *is constrained optimal, then* η^{π^*} *is an optimal solution to LP* (3.9).

(c) *If, in addition,* $U \neq \varnothing$ *and Assumptions* C(1) *and* C(2) *are satisfied, then an optimal solution* η^* *to LP* (3.9) *exists, and the policy* ϕ^{η^*} *is constrained optimal.*

The proof of Theorem 3.13 appears in Section 5.

In particular, when S and $A(x)$ are finite, then LP (3.9) is the form of

$$\text{minimize} \sum_{x\in S}\sum_{a\in A(x)} \frac{1}{\alpha} c_0(x,a)\eta(x,a)$$

$$\text{subject to} \begin{cases} \displaystyle\sum_{x\in S}\sum_{a\in A(x)} c_1(x,a)\eta(x,a) \le \alpha d_1, \\ \qquad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \displaystyle\sum_{x\in S}\sum_{a\in A(x)} c_n(x,a)\eta(x,a) \le \alpha d_N, \\ \displaystyle\alpha\sum_{a\in A(x)} \eta(x,a) = \alpha\gamma(x) + \sum_{y\in S}\sum_{a\in A(y)} q(x|y,a)\eta(y,a), \\ \qquad \forall x \in S, \eta(x,a) \ge 0, x \in S, a \in A(x), \end{cases} \tag{3.10}$$

which is a LP and can be solved by many methods such as the well-known simplex method.

To state the structure of constrained optimal policies, we need to recall some concepts. We say that under $\phi \in \Pi_s$, there are $m(x,\phi)$ randomizations at $x \in S$ if there are $m(x,\phi)+1$ actions $a \in A(x)$ for which $\phi(a|x) > 0$. When S and $A(x)$ are finite, we call $\#(\phi) := \sum_{x\in S} m(x,\phi)$ the number of randomizations under ϕ.

Thus, following Theorem 3.8 in [1] and Theorem 3.13 above, we have the following fact.

COROLLARY 3.14. *Suppose that S and $A(x)$ are finite. Let η^* be an optimal basic solution to LP (3.10). Then, the policy ϕ^{η^*} is constrained optimal, where ϕ^{η^*} is given by*

$$\phi^{\eta^*}(a|x) = \begin{cases} \dfrac{\eta^*(x,a)}{\hat{\eta}^*(x)}, & \text{when } \hat{\eta}^*(x) := \displaystyle\sum_{a\in A(x)} \eta^*(x,a) > 0 \\ & \text{and } a \in A(x), \\ I_{\{a(x)\}}(a), & \text{when } \hat{\eta}^*(x) = 0 \text{ and } a \in A(x), \end{cases} \tag{3.11}$$

for all $x \in S$, $a(x) \in A(x)$ is chosen arbitrarily. Further, $\#(\phi^{\eta^*}) \le N$.

Corollary 3.14 provides the structure of a constrained optimal policy for finite S and $A(x)$, and it is proven for the case of denumerable states and a single constraint in [13, 42]. For a more general case of Polish spaces, we have the following facts, in which the first one (i.e., Theorem 3.15) establishes the relationship between stationary policies in F and extreme points in $\mathcal{M}_o$, and the second one (i.e., Theorem 3.16) shows a structure of a constrained optimal policy.

THEOREM 3.15. *Suppose that Assumptions A, B(2), B(3) and C(3) hold. Then*:

(a) *η^f is an extreme point in $\mathcal{M}_o$ for each $f \in F$.*

(b) *If, for each $\phi \in \Pi_s$ and $D \in \mathcal{B}(S)$ with $\hat{\eta}^\phi(D) > 0$, there exists state $x \in D$ (depending on D and ϕ) such that $\hat{\eta}^\phi(\{x\}) > 0$, then η is an extreme point in $\mathcal{M}_o$ if and only if there exists a policy $f \in F$ such that $\eta = \eta^f$.*

[The condition in Theorem 3.15(b) is satisfied when S is denumerable.]

The proof Theorem 3.15 appears in Section 5.

THEOREM 3.16. *Suppose that Assumptions A, B, C and the conditions for Theorem 3.15(b) are satisfied. Then, there exists a constrained optimal policy $\pi^* \in \Pi_s$, which is a mixture of $(N+1)$ stationary policies, that is, there exists $(N+1)$ numbers $p_n \ge 0$ and policies $f_n \in F$ $(1 \le n \le N+1)$ such that $\pi^* = \phi^{(p_1\eta^{f_1}+\cdots+p_{N+1}\eta^{f_{N+1}})}$ and $p_1+\cdots+p_{N+1} = 1$.*

The proof of Theorem 3.16 appears in Section 5.

REMARK 3.17. The arguments of Theorems 3.11, 3.13, 3.15 and 3.16 do not depend on the data in model (2.1), but they are based on Theorem 3.5. Thus, the discrete-time versions of Theorems 3.11, 3.13, 3.15 and 3.16 are still true because Theorem 3.5 is established in [1, 2, 21] for discrete-time MDPs.

4. Examples. In this section, we illustrate our conditions and main results with examples.

EXAMPLE 4.1. Let $S := (-\infty, \infty)$, $A(x) := [\beta_0, \beta(|x| + 1)]$ for each $x \in S$ with some constants $0 < \beta_0 < \beta$. Suppose that the reward $r(x, a)$ and costs $c_n(x, a)$ $(1 \le n \le N)$ are given. We consider the transition rates $q(\cdot|x, a)$ given by

$$q(D|x, a) := (|x| + 1)\left[\int_{D-\{x\}} f(y|x, a)\, dy - \delta_x(D)\right] \quad \text{for } (x, a) \in K,\ D \in \mathcal{B}(S), \tag{4.1}$$

where $f(y|x, a) := \frac{1}{\sqrt{2\pi a}} e^{-(y-x)^2/(2a)}$ is the density function of Gaussian distribution $N(x, a)$.

We now aim to find conditions that ensure the existence of constrained optimal policies for Example 4.1. To do so, we need the following hypotheses.

ASSUMPTION D. Let α, γ, d_n and U $(\neq\varnothing)$ be as in (2.11).

(1) $\alpha > 6\beta$ and $\int_S x^4 \gamma(dx) < \infty$ (hence, there exists a constant ρ such that $6\beta < \rho < \alpha$);

(2) $c_n(x, a)$ $(0 \le n \le N)$ are continuous on K and $|c_n(x, a)| \le L'(x^2 + 1)$ for all $(x, a) \in K$, with some constant $L' > 0$, where $c_0(x, a) := -r(x, a)$.

Then, we have the following result.

PROPOSITION 4.2. *Under Assumption D, Example 4.1 satisfies Assumptions A, B and C. Therefore (by Theorem 3.11), there exists a constrained optimal policy for Example 4.1.*

PROOF. For each $m \ge 1$ and $x \in S$, let

$$S_m := [-m, m], \qquad K_m := \{(x, a) | x \in S_m, a \in A(x)\},$$
$$w'(x) := x^2 + 1, \qquad w(x) := x^4 + 1. \tag{4.2}$$

To verify Assumption A, it suffices to verify Assumption A(1) because Assumptions A(2) and A(3) follow from (4.2) and (4.1). Indeed, by (4.1) and a straightforward calculation, we have

$$\begin{aligned}\int_S w(y)q(dy|x,a) &= 6(x^2a+3a^2)(|x|+1)\\ &\le \beta w(x)+b \qquad \text{for some constant } b>0,\end{aligned} \tag{4.3}$$

which implies Assumption A(1).

Obviously, Assumption B follows from (4.3) and Assumptions D(1) and D(2).

To verify Assumption C, for any $g \in C_b(S)$, by (4.1) we have the following:

$$\int_S g(y)q(dy|x,a) = (|x|+1)\left[\int_{-\infty}^{\infty} g(y)\frac{1}{\sqrt{2\pi a}}e^{-(y-x)^2/(2a)}\,dy - g(x)\right],$$

which, together with the dominated convergence theorem, implies Assumption C(1). Therefore, Assumption C holds because Assumptions C(2) and C(3) follow from (4.1) and (4.2).

Using Example 4.1, we present computable examples for unconstrained optimal policies.

EXAMPLE 4.3. With the same data as in Example 4.1, we further suppose that $r(x,a)$ in Example 4.1 is given by

$$r(x,a) := px^2 - \delta a^2 \qquad \text{for } (x,a) \in K, \tag{4.4}$$

where $p, \delta > 0$ are fixed constants.

ASSUMPTION E. Let β_0 and β be as in Example 4.1, and L' as in Assumption D(2).

(1) $d_n \ge L'[\alpha \int_S x^4\gamma(dx) + \alpha + b]/[\alpha(\alpha-\beta)]$ for all $1 \le n \le N$), with $b := \beta(\frac{\rho+2\beta}{\rho-\beta}+2)^2$;

(2) $2\alpha\beta_0 - \beta_0^2 \le \frac{p}{\delta} \le \min\{\alpha^2, 2\alpha\beta - \beta^2\}$, with p, δ as in (4.4).

PROPOSITION 4.4. *Suppose that Assumptions* D *and* E *hold. Then*:

(a) *Example* 4.3 *satisfies Assumptions* A, B *and* C. *Moreover*, $V_r(U) = \int_S u(x)\gamma(dx)$, *where*

$$\begin{aligned}u(x) = {}& \big(2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta}\big)x^2 + \left(4\delta\alpha - 4\sqrt{\delta^2\alpha^2 - p\delta} - \frac{2p}{\alpha}\right)|x|\\ &+ 2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta} - \frac{p}{\alpha}.\end{aligned}$$

(b) *The stationary policy f^* is unconstrained optimal for Example 4.3, where*

$$f^*(x) := \left(\alpha - \sqrt{\alpha^2 - \frac{p}{\delta}}\right)(|x| + 1) \qquad \forall x \in S.$$

PROOF. Note that Assumptions E(1) and D imply that $U = \Pi$ (by Theorem 3.3), and so the problem (2.12) becomes an unconstrained optimality problem. Thus, as in Proposition 4.2, under Assumptions D and E, we see that all assumptions in Theorem 3.3 in [12] are satisfied. Hence, Theorem 3.3 in [12] ensures the existence of a function u in $B_w(S)$ such that, for each $x \in S$ and $\pi \in \Pi$,

$$\alpha u(x) = \sup_{a \in A(x)} \left\{ r(x,a) + \int_S u(y) q(dy|x,a) \right\} \quad \text{and} \quad u(x) \geq V_r(x,\pi). \tag{4.5}$$

To obtain the analytic expression of u, we assume for a moment that

$$u(x) := l_2 x^2 + l_1 x + l_0 \qquad \text{for } x \in S, \text{ with some constants } l_1, l_2, l_2. \tag{4.6}$$

Then, using (4.1), (4.4) and (4.5), by a straightforward calculation we have

$$\alpha(l_2 x^2 + l_1 x + l_0) = \sup_{a \in A(x)} \left\{ px^2 - \delta\left(a - \frac{l_2(|x|+1)}{2\delta}\right)^2 + \frac{l_2^2(|x|+1)^2}{4\delta} \right\}, \tag{4.7}$$

which implies that $f^*(x) := \frac{l_2(|x|+1)}{2\delta}$ attains the maximum of the right-hand side of (4.7). Therefore, by Theorem 3.3 in [12], we have

$$V_r(x, f^*) = u(x) \quad \text{and} \quad \alpha(l_2 x^2 + l_1 x + l_0) = px^2 + \frac{l_2^2(|x|+1)^2}{4\delta} \qquad \forall x \in S. \tag{4.8}$$

Comparing with the coefficients of both sides in (4.8), we obtain

$$\alpha l_2 = p + \frac{l_2^2}{4\delta}, \qquad \alpha l_1 = \begin{cases} \dfrac{l_2^2}{2\delta}, & \text{if } x \geq 0, \\ -\dfrac{l_2^2}{2\delta}, & \text{otherwise,} \end{cases} \qquad \alpha l_0 = \frac{l_2^2}{4\delta}. \tag{4.9}$$

Under Assumption E, solving the system of equations (4.9) gives

$$l_2 = 2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta}, \qquad l_0 = 2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta} - \frac{p}{\alpha},$$

$$l_1 = \begin{cases} 4\delta\alpha - 4\sqrt{\delta^2\alpha^2 - p\delta} - \dfrac{2p}{\alpha}, & \text{if } x \geq 0, \\ -\left(4\delta\alpha - 4\sqrt{\delta^2\alpha^2 - p\delta} - \dfrac{2p}{\alpha}\right), & \text{otherwise,} \end{cases}$$

which, together with (4.6) and (4.8), yields

$$u(x) = \bigl(2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta}\bigr)x^2 + \biggl(4\delta\alpha - 4\sqrt{\delta^2\alpha^2 - p\delta} - \frac{2p}{\alpha}\biggr)|x| + 2\delta\alpha - 2\sqrt{\delta^2\alpha^2 - p\delta} - \frac{p}{\alpha},$$

$$f^*(x) = \biggl(\alpha - \sqrt{\alpha^2 - \frac{p}{\delta}}\biggr)(|x| + 1) \in A(x) \quad \text{and} \quad V_r(x, f^*) = u(x) \qquad \forall x \in S.$$

This, together with (4.5) and (2.10), completes the proof of this proposition. □

EXAMPLE 4.5. Let $S := (-\infty, \infty)$, $A(x) := [0, \beta(|x| + 1)]$ for each $x \in S$ with some constant $\beta > 0$, and the reward $r(x, a)$ and transition rates $q(\cdot|x, a)$ are defined as follows: for each $(x, a) \in K$ and $D \in \mathcal{B}(S)$,

$$q(D|x, a) := (\beta|x| + a)\biggl[\int_{D-\{x\}} \frac{1}{\sqrt{2\pi(\beta(|x| + 1) - a + 1)}} \times e^{-(y-x)^2/(2(\beta(|x|+1)-a+1))}\, dy - \delta_x(D)\biggr].$$

$$r(x, a) := p|x|a - \delta a^2 \qquad \text{for } (x, a) \in K, \text{ with } p, \delta > 0.$$

ASSUMPTION E. $\alpha > \beta^2$; $\int_S x^2\gamma(dx) < \infty$; and $\beta \geq \max\{1, \frac{p}{2\delta}\}$.

Then as the arguments for Example 4.3 in Proposition 4.4, we have the following results.

PROPOSITION 4.6. *Under Assumption* E, *Example* 4.5 *satisfies Assumptions* A, B *and* C. *Moreover, if, in addition,* $U = \Pi$, *then* $V_r(U) = \int_S u(x)\gamma(dx)$, *where*

$$u(x) = \frac{1}{2}\delta\bigl(\sqrt{\kappa + 1} - 1\bigr)x^2 + \frac{1}{2\alpha\kappa}\bigl[p\bigl(\sqrt{\kappa + 1} - 1\bigr) + \kappa\delta\beta\bigr](\beta + 1)\bigl(\sqrt{\kappa + 1} - 1\bigr)|x| + \frac{1}{8\alpha\kappa}\delta(\beta + 1)^2\bigl(\sqrt{\kappa + 1} - 1\bigr)^3$$

with $\kappa := \frac{p^2}{\delta^2(\alpha - \beta^2)} > 0$, *and the following stationary policy* f^* *is unconstrained optimal*:

$$f^*(x) := \frac{p(\sqrt{\kappa + 1} - 1)}{\delta\kappa}|x| + \frac{1}{2\kappa}(\beta + 1)\bigl(\sqrt{\kappa + 1} - 1\bigr)^2 \qquad \forall x \in S.$$

PROOF. The proof of Proposition 4.6 is similar to that of Proposition 4.2, and thus the details are omitted here. □

REMARK 4.7. In Examples 4.1, 4.3 and 4.5, the transition rates are *unbounded*, and the reward and costs are allowed to be *unbounded from above and from below*. In contrast, the transition rates in [18, 26, 27, 30, 37, 39, 40] are assumed to be *bounded*, and the costs in [11, 19, 20, 22, 27, 29] are assumed to be *nonnegative*. Moreover, Examples 4.3 and 4.5 seem to be first computable examples for the unconstrained optimal policies for discounted continuous-time MDPs in Polish spaces.

5. Proofs of the main results. In this section, we give proofs of Theorems 3.1, 3.5, 3.11, 3.13, 3.15, 3.16 and of Lemmas 3.8 and 3.9, which are stated in Section 3.

To prove Theorems 3.1, we need the following two lemmas.

LEMMA 5.1. *Suppose that real-valued measurable functions $\bar{w} \geq 0$ on S and $\bar{q}_t(D|x)$ on $\mathbb{R}^0_+ \times \mathcal{B}(S) \times S$ satisfy the following*: *for each $t \geq 0$, $D \in \mathcal{B}(S)$ and $x \in S$*:

(1) *$\bar{q}_t(\cdot|x)$ is a signed measure on $\mathcal{B}(S)$ such that $\bar{q}_t(S|x) \equiv 0$, $\bar{q}_t(D|x) \geq 0$ for all $x \notin D$ and $\bar{q}_t(x) := \bar{q}_t(S - \{x\}|x) < \infty$;*

(2) *$\int_S \bar{w}(y)\bar{q}_t(dy|x) \leq \bar{\rho}\bar{w}(x) + \bar{b}$, with constants $\bar{\rho} \neq 0$ and $\bar{b} \geq 0$.*

Then nonnegative function

$$\bar{h}(s,x,t) := e^{\bar{\rho}(t-s)}\bar{w}(x) + \frac{\bar{b}}{\bar{\rho}}\big(e^{\bar{\rho}(t-s)} - 1\big) \tag{5.1}$$

satisfies the following inequality:

$$\int_s^t \int_{S-\{x\}} e^{-\int_s^z \bar{q}_v(x)\,dv}\bar{q}_z(dy|x)\bar{h}(z,y,t)\,dz + e^{-\int_s^t \bar{q}_v(x)\,dv}\bar{w}(x) \leq \bar{h}(s,x,t)$$

for all $x \in S$ and $0 \leq s \leq t < \infty$.

PROOF. Under conditions (1) and (2), a straightforward calculation gives

$$\begin{aligned}
&\int_s^t \int_{S-\{x\}} e^{-\int_s^z \bar{q}_v(x)\,dv}\bar{q}_z(dy|x)\bar{h}(z,y,t)\,dz \\
&\quad \leq \int_s^t e^{-\int_s^z \bar{q}_v(x)\,dv}\bigg[e^{\bar{\rho}(t-z)}\bigg(\bar{\rho}\bar{w}(x) + \bar{b} \\
&\qquad\qquad + \bar{w}(x)\bar{q}_z(x) + \frac{\bar{b}}{\bar{\rho}}\bar{q}_z(x)\bigg) - \frac{\bar{b}}{\bar{\rho}}\bar{q}_z(x)\bigg]\,dz \\
&\quad = \bar{h}(s,x,t) - e^{-\int_s^t \bar{q}_v(x)\,dv}\bar{w}(x),
\end{aligned}$$

which verifies this lemma. □

LEMMA 5.2. *Suppose that Assumption* A(1) *holds for* $\rho \neq 0$. *Then, for any* $\pi \in \Pi$ *and* $x \in S$,

$$E_x^\pi\big[w(\xi_t)I_{\{t<T_{k+1}\}}\big] \le e^{\rho t}w(x) + \frac{b}{\rho}(e^{\rho t}-1) \qquad \forall k \ge 0 \text{ and } t \ge 0,$$

where w *and* b *are from Assumption* A(1).

PROOF. Fix any $\pi \in \Pi$, $l \ge 1$, and $(x_0, \theta_1, x_1, \dots, x_{l-1}, \theta_l) \in (S \times \mathbb{R}_+^0)^l$. Let $m_l(\cdot|h_l, t)$ be as in (2.8). Then, it follows from Assumption A(1) that the following function on $\mathbb{R}_+^0 \times \mathcal{B}(S) \times S$:

$$\bar{q}_t(D|x) := \begin{cases} m_l(D|x_0, \theta_1, x_1, \dots, \theta_l, x, t), & \text{if } x \notin D, \\ -m_l(S|x_0, \theta_1, x_1, \dots, \theta_l, x, t), & \text{if } D = \{x\}, \end{cases}$$

satisfies conditions (1) and (2) for Lemma 5.1.

Let $h(s,x,t) := e^{\rho(t-s)}w(x) + \frac{b}{\rho}(e^{\rho(t-s)}-1)$ for all $x \in S$ and $t \ge s \ge 0$. Then, for each fixed $x \in S$ and $0 \le s \le t$, by Lemma 5.1 we have

$$\begin{aligned} &\int_s^t \int_{S-\{x\}} m_l(dy|h_{l-1}, \theta_l, x, z - T_l)h(z, y, t) \\ &\qquad \times e^{-\int_s^z m_l(S|h_{l-1},\theta_l,x,v-T_l)\,dv}\,dz \\ &\quad + w(x)e^{-\int_s^t m_l(S|h_{l-1},\theta_l,x,v-T_l)\,dv} \\ &= \int_{s-T_l}^{t-T_l} \int_{S-\{x\}} m_l(dy|h_{l-1}, \theta_l, x, \tilde{u})h(\tilde{u}, y, t - T_l) \\ &\qquad \times e^{-\int_{s-T_l}^{\tilde{u}} m_l(S|h_{l-1},\theta_l,x,\tilde{v})\,d\tilde{v}}\,d\tilde{u} \\ &\quad + w(x)e^{-\int_{s-T_l}^{t-T_l} m_l(S|h_{l-1},\theta_l,x,\tilde{v})\,d\tilde{v}} \\ &\le h(s - T_l, x, t - T_l) = h(0, x, t - s). \end{aligned} \tag{5.2}$$

Moreover, by (2.5) and (2.9), we have

$$\begin{aligned} &E_x^\pi\big[w(\xi_t)I_{\{t<T_{k+1}\}}|\mathcal{F}_{T_k}\big] \\ &\quad = e^{-\int_0^{t-T_k} m_k(S|h_k,v)\,dv}w(x_k)I_{\{T_k \le t\}} + I_{\{T_k>t\}}\sum_{m=1}^k I_{\{T_{m-1}\le t<T_m\}}w(x_{m-1}). \end{aligned}$$

Now, using (5.2) at $l = k$, $s = T_k = T_l$, $x = x_k = x_l$, gives

$$\begin{aligned} &E_x^\pi\big[w(\xi_t)I_{\{t<T_{k+1}\}}|\mathcal{F}_{T_k}\big] \\ &\quad \le I_{\{T_k\le t\}}h(T_k, x_k, t) + I_{\{T_k>t\}}\sum_{m=1}^k I_{\{T_{m-1}\le t<T_m\}}w(x_{m-1}), \end{aligned}$$

which implies that the following (5.3) holds for $n = 0$:

$$E_x^\pi\big[w(\xi_t)I_{\{t<T_{k+1}\}}|\mathcal{F}_{T_{k-n}}\big] \le I_{\{T_{k-n}\le t\}}h(T_{k-n},x_{k-n},t) + I_{\{T_{k-n}>t\}}\sum_{m=1}^{k-n} I_{\{T_{m-1}\le t<T_m\}}w(x_{m-1}) \quad \forall k\ge n\ge 0. \tag{5.3}$$

Suppose that (5.3) holds for some $0 \le n < k$. Then, by (2.9) we have

$$\begin{aligned}
&E_x^\pi\big[w(\xi_t)I_{\{t<T_{k+1}\}}|\mathcal{F}_{T_{k-n-1}}\big]\\
&\quad\le E_x^\pi\Bigg[I_{\{T_{k-n}\le t\}}h(T_{k-n},x_{k-n},t)\\
&\qquad\qquad + I_{\{T_{k-n}>t\}}\sum_{m=1}^{k-n} I_{\{T_{m-1}\le t<T_m\}}w(x_{m-1})\Big|\mathcal{F}_{T_{k-n-1}}\Bigg]\\
&\quad= E_x^\pi\big[I_{\{T_{k-n}\le t\}}h(T_{k-n},x_{k-n},t)\\
&\qquad\qquad + I_{\{T_{k-n}>t\}}I_{\{T_{k-n-1}\le t<T_{k-n}\}}w(x_{k-n-1})|\mathcal{F}_{T_{k-n-1}}\big]\\
&\qquad + I_{\{T_{k-n}>t\}}\sum_{m=1}^{k-n-1} I_{\{T_{m-1}\le t<T_m\}}w(x_{m-1})\\
&\quad= I_{\{T_{k-n-1}\le t\}}\Bigg[\int_0^{t-T_{k-n-1}}\int_{S-\{x_{k-n-1}\}} m_{k-n-1}(dy|h_{k-n-1},\tilde t)\\
&\qquad\qquad\qquad\qquad\qquad\qquad \times h(T_{k-n-1}+\tilde t,y,t)\\
&\qquad\qquad\qquad\qquad\qquad\qquad \times e^{-\int_0^{\tilde t} m_{k-n-1}(S|h_{k-n-1},\tilde v)\,d\tilde v}\,d\tilde t\\
&\qquad\qquad\qquad + e^{-\int_0^{t-T_{k-n-1}} m_{k-n-1}(S|h_{k-n-1},\tilde v)\,d\tilde v}w(x_{k-n-1})\Bigg]\\
&\qquad + I_{\{T_{k-n-1}>t\}}\sum_{m=1}^{k-n-1} I_{\{T_{m-1}\le t<T_m\}}w(x_{m-1}),
\end{aligned}$$

which together with $h(T_{k-n-1}+\tilde t,y,t) = h(\tilde t,y,t-T_{k-n-1})$ and (5.2) again, gives

$$\begin{aligned}
&E_x^\pi\big[w(\xi_t)I_{\{t<T_{k+1}\}}|\mathcal{F}_{T_{k-n-1}}\big]\\
&\quad\le I_{\{T_{k-n-1}\le t\}}h(T_{k-n-1},x_{k-n-1},t)\\
&\qquad + I_{\{T_{k-n-1}>t\}}\sum_{m=1}^{k-n-1} I_{\{T_{m-1}\le t<T_m\}}w(x_{k-1}).
\end{aligned}$$

Hence, (5.3) holds for all $0 \le n \le k$, and so this lemma follows from (5.3) at $n = k$. □

PROOF OF THEOREM 3.1. (a) We first prove the following fact:

$$P_x^\pi\big(\xi_t I_{\{T_k\le t<T_{k+1}\}}\notin S_l\text{: for some } k\ge 0\big)\to 0 \qquad \text{as } l\to\infty. \tag{5.4}$$

To prove (5.4), let $\Gamma_l := \{e : \xi_t(e) I_{\{T_k\le t<T_{k+1}\}}(e)\notin S_l \text{ for some } k\ge 0\}$ for any $l\ge 1$.

Suppose that, for some $\varepsilon>0$ and any $\tilde L\ge 1$, there exists $l>\tilde L$ such that

$$P_x^\pi(\Gamma_l) = P_x^\pi\big(\{e : \xi_t(e) I_{\{T_k\le t<T_{k+1}\}}(e)\notin S_l \text{ for some } k\ge 0\}\big) > \varepsilon. \tag{5.5}$$

Then, by Assumption A(2), we can take the corresponding l such that (5.5) holds and also the following inequality:

$$w(y) > \left[e^{\tilde\rho t}w(x) + \frac{b}{\tilde\rho}(e^{\tilde\rho t}-1)\right]\Big/\varepsilon \qquad \forall y\notin S_l, \tag{5.6}$$

is satisfied, where $\tilde\rho := |\rho|+1$.

For the taken $l\ge 1$ in (5.6), let us define new transition rates $\tilde q(D|x,a)$ as follows:

$$\tilde q(D|x,a) := \begin{cases} q(D|x,a), & \text{if } x\in S_l,\\ 0, & \text{if } x\notin S_l, \end{cases} \qquad \text{for } (x,a)\in K.$$

The quantities such as probabilities corresponding to $\tilde q(D|x,a)$ are equipped with the tilde.

We next to prove that

$$\begin{aligned} &P_x^\pi\big(\xi_t I_{\{T_k\le t<T_{k+1}\}}\in S_l \text{ for all } k\ge 0\big)\\ &\quad = \tilde P_x^\pi\big(\xi_t I_{\{T_k\le t<T_{k+1}\}}\in S_l \text{ for all } k\ge 0\big). \end{aligned} \tag{5.7}$$

Indeed, it is obvious that

$$P_x^\pi(X_0\in S_l) = \tilde P_x^\pi(X_0\in S_l) = I_{S_l}(x).$$

Let $X_k^t := X_k I_{\{T_k\le t<T_{k+1}\}}$. Then, by (2.5) we have $\{\xi_t I_{\{T_k\le t<T_{k+1}\}}\in S_l\} = \{X_k^t\in S_l\}$. We now suppose that for some $n\ge 0$,

$$\begin{aligned} &P_x^\pi(\{X_k^t\in S_l, 0\le k\le n\}\cap\Gamma)\\ &\quad = \tilde P_x^\pi(\{X_k^t\in S_l, 0\le k\le n\}\cap\Gamma) \qquad \forall\Gamma\in\mathcal{B}(\hat H_n), \end{aligned} \tag{5.8}$$

where P_x^π and $\tilde P_x^\pi$ are regarded as the marginal on $\hat H_{n+1}$.

Using the notation in (2.8) and (2.9), for any $D\in\mathcal{B}(S)$, $0<t_1<t_2<\infty$, we have

$$\begin{aligned} &P_x^\pi\big(\{X_k^t\in S_l, 0\le k\le n, \text{ and } X_{n+1}^t\in S_l\}\cap\{\Gamma\times(t_1,t_2)\times D\}\big)\\ &\quad = \int_{t_1}^{t_2}\int_\Gamma P_x^\pi(dh_n) I_{\{X_k^t\in S_l, 0\le k\le n\}} I_{\{X_{n+1}^t\in S_l\cap D\}}\\ &\qquad\qquad \times m_n(S_l\cap D|h_n,\tilde t\,) e^{-\int_0^{\tilde t} m_n(S|h_n,v)\,dv}\,d\tilde t \end{aligned}$$

$$= \int_{t_1}^{t_2} \int_{\Gamma} \tilde{P}_x^{\pi}(dh_n) I_{\{X_k^t \in S_l, 0 \le k \le n\}} I_{\{X_{n+1}^t \in S_l \cap D\}}$$

$$\times \tilde{m}_n(S_l \cap D | h_n, \tilde{t}) e^{-\int_0^{\tilde{t}} \tilde{m}_n(S|h_n, v)\, dv}\, d\tilde{t}$$

$$= \tilde{P}_x^{\pi}\big(\{X_k^t \in S_l, 0 \le k \le n, \text{and} X_{n+1}^t \in S_l\} \cap \{\Gamma \times (t_1, t_2) \times D\}\big),$$

which together with the arbitrariness of $D \in \mathcal{B}(S)$ and $0 \le t_1 < t_2$ implies (5.8) for $n+1$, and thus (5.7) follows from the induction.

Thus, from (5.5) and (5.7), we have

$$\tilde{P}_x^{\pi}(\Gamma_l) = \tilde{P}_x^{\pi}\big(\xi_t I_{\{T_k \le t < T_{k+1}\}} \notin S_l \text{ for some } k \ge 0\big) > \varepsilon. \tag{5.9}$$

Moreover, since $\|\tilde{q}\| := \sup_{x \in S, a \in A(x)} |\tilde{q}(\{x\}|x,a)| = \sup_{x \in S_l, a \in A(x)} |q(\{x\}|x, a)| < \infty$, we now show by induction that

$$\tilde{E}_x^{\pi}[e^{-T_k}] \le \big[1 - e^{-\|\tilde{q}\|}(1 - e^{-1})\big]^k \qquad \forall k \ge 1. \tag{5.10}$$

In fact, by (2.8) we have $|\tilde{m}_k(S|h_k) \le \|\tilde{q}\|$ for all $k \ge 1$, and it follows from (2.9) that

$$\begin{aligned} \tilde{E}_x^{\pi}[e^{-T_1}] &= \int_0^1 \tilde{m}_0(S|x) e^{-\tilde{m}_0(S|x)t} e^{-t}\, dt + \int_1^{\infty} \tilde{m}_0(S|x) e^{-\tilde{m}_0(S|x)t} e^{-t}\, dt \\ &\le 1 - e^{-\|\tilde{q}\|} \int_0^1 e^{-t}\, dt = \big[1 - e^{-\|\tilde{q}\|}(1 - e^{-1})\big]. \end{aligned} \tag{5.11}$$

Suppose that (5.10) holds for some $k \ge 1$. Then, as the arguments of (5.11), from (2.8) and (2.9) we also have $\tilde{E}_x^{\pi}[e^{-T_{k+1}}] \le \tilde{E}_x^{\pi}[e^{-T_k}[1 - e^{-\|\tilde{q}\|}(1 - e^{-1})]] \le [1 - e^{-\|\tilde{q}\|}(1 - e^{-1})]^{k+1}$, and so (5.10) follows. Hence, by (5.10) and the Chebychev inequality we have

$$\begin{aligned} \tilde{P}_x^{\pi}(T_\infty \le t) \le \tilde{P}_x^{\pi}(T_k \le t) &= \tilde{P}_x^{\pi}(e^{-T_k} \ge e^{-t}) \le e^t \tilde{E}_x^{\pi}[e^{-T_k}] \\ &\le e^t \big[1 - e^{-\|\tilde{q}\|}(1 - e^{-1})\big]^k \end{aligned}$$

for all $k \ge 1$, and so $\tilde{P}_x^{\pi}(T_\infty \ge t) = 1$. Since $t > 0$ can be arbitrary, we have $\tilde{P}_x^{\pi}(T_\infty = \infty) = 1$, and therefore, $\sum_{k=0}^{\infty} \tilde{P}_x^{\pi}(T_k \le t < T_{k+1}) = 1$. Since Assumption A(1) still holds when ρ and $q(D|x,a)$ are replaced with $\bar{\rho}$ and $\tilde{q}(D|x,a)$, respectively, by Lemma 5.2 we have

$$\tilde{E}_x^{\pi}[w(\xi_t)] = \lim_{k \to \infty} \tilde{E}_x^{\pi}\big[w(\xi_t) I_{\{t < T_{k+1}\}}\big] \le e^{\tilde{\rho} t} w(x) + \frac{b}{\tilde{\rho}}(e^{\tilde{\rho} t} - 1). \tag{5.12}$$

On the other hand, using (5.6) and (5.9), we see

$$\begin{aligned} \tilde{E}_x^{\pi}[w(\xi_t)] &= \tilde{E}_x^{\pi}[w(\xi_t)|\Gamma_l] \tilde{P}_x^{\pi}(\Gamma_l) + \tilde{E}_x^{\pi}[w(\xi_t)|\Gamma_l^c] \tilde{P}_x^{\pi}(\Gamma_l^c) \\ &> e^{\tilde{\rho} t} w(x) + \frac{b}{\tilde{\rho}}(e^{\tilde{\rho} t} - 1), \end{aligned}$$

which contradicts to (5.12), and thus (5.4) is proved.

Since $\Gamma_{l+1} \subseteq \Gamma_l$ for all $l \geq 1$, by (5.4) we conclude that $P_x^\pi(\bigcap_{l\geq 0}\Gamma_l) = 0$, and so

$$P_x^\pi\big(\{\text{for each } l \geq 1, \text{ there exists } k \text{ such that } \xi_t I_{\{T_k \leq t < T_{k+1}\}} \notin S_l\}\big) = 0. \tag{5.13}$$

Since $\{\inf\{s : \xi_s \notin S_l\} \leq t\} \subseteq \{\xi_t I_{\{T_k\leq t<T_{k+1}\}} \notin S_l, \text{ for some } k \geq 1\}$, by (5.13) we conclude $P_x^\pi(\inf\{s : \xi_s \notin S_l\} \leq t, l = 1, \ldots) = 0$, and thus $P_x^\pi(\inf\{s : \xi_s \notin S_l\} > t$, for some $l \geq 1) = 1$, or, equivalently, $P_x^\pi(\xi_s \in S_l$ for all $s \in [0, t]$, for some $l \geq 1) = 1$. For any $k \geq 1$, let $B_k := \{\xi_s \in S_l$ for all $s \in [0, k]$, for some $l \geq 1\}$. Then, $B_{k+1} \subseteq B_k$ and $P_x^\pi(B_k) = 1$ for all $k \geq 1$, and thus $P_x^\pi(\bigcap_{k=1}^\infty B_k) = 1$, which together with (2.5) implies $P_x^\pi(T_\infty = \infty) = 1$. To further prove $P_x^\pi(\xi_t \in S) = 1$, using the facts $\sum_{k\geq 0} P_x^\pi(T_k \leq t < T_{k+1}) = P_x^\pi(T_\infty = \infty) = 1$ and $P_x^\pi(\xi_t \in S|T_k \leq t < T_{k+1}) = 1$ for all $k \geq 1$, we have that $P_x^\pi(\xi_t \in S) = \sum_{k\geq 0} P_x^\pi(\xi_t \in S|T_k \leq t < T_{k+1}) P_x^\pi(T_k \leq t < T_{k+1}) = 1$, and thus (a) follows.

(b) First, consider the case of $\rho \neq 0$. Since $\sum_{k=0}^\infty P_x^\pi(T_k \leq t < T_{k+1}) = 1$ for all $t \geq 0$,

$$E_x^\pi[w(\xi_t)] = E_i^\pi\left[w(\xi_t)\sum_{k=0}^\infty I\{T_k \leq t < T_{k+1}\}\right] = \lim_{k\to\infty} E_i^\pi[w(\xi_t)I\{t < T_{k+1}\}],$$

which together with Lemma 5.2 implies the first part of (b). Moreover, the results for the case of $\rho = 0$ can be obtained by letting $\rho \downarrow 0$.

(c) Define an integer-valued random measure $\tilde{\mu}^*$ on $\mathcal{B}(\mathbb{R}_+^0) \times \mathcal{B}(S)$

$$\tilde{\mu}^*(dt, dx) := \sum_{k\geq 1} I_{\{T_k<\infty\}}\delta_{(T_k, X_{k-1})}(dt, dx), \tag{5.14}$$

which counts the exits from dx. Then, as Lemma 4.28 in [28], the random measure

$$\tilde{\nu}^\pi(e, dt, dx) := -\left[\int_A \pi(da|e, t)q(dx|\xi_{t-}(e), a)I_{dx}(\xi_{t-}(e))\right]dt$$

is a dual predictable projection of the measure $\tilde{\mu}^*$ with respect to $\mathcal{P}$ and P_γ^π (for any fixed policy $\pi \in \Pi$ and initial distribution γ). Hence, by (4.5) in [28] we have

$$\begin{aligned} E_x^\pi[\tilde{\mu}^*((0, t], D)] &= E_x^\pi[\tilde{\nu}^\pi((0, t], D)] \\ &\leq E_x^\pi\left[\int_0^t\int_A \pi(da|e, s)\sup_{x\in D} q^*(x)\,ds\right] < \infty \qquad \forall t \geq 0, \end{aligned}$$

which together with $|\mu^*((0, t], D) - \tilde{\mu}^*((0, t], D)| \leq 1$ and (4.5) in [28] again, implies

$$E_x^\pi[\mu^*((0, t], D)] = E_x^\pi[\nu^\pi((0, t], D)] < \infty.$$

Thus, using the obvious representation $I_{\{\xi_t\in D\}} = I_D(x) + \mu^*((0, t], D) - \tilde{\mu}^*((0, t], D)$, by taking the expectation E_x^π of the representation we see that (c) is true. □

PROOF OF THEOREM 3.5. (a) For the given D, by Theorem 3.1(c) and (3.1) we have

$$\begin{aligned}\hat{\eta}^{\pi}(D) &= \gamma(D) + \alpha \int_0^{\infty} e^{-\alpha t} E_{\gamma}^{\pi}\left[\int_0^t \int_A \pi(da|e,s) q(D|\xi_{s-}(e),a)\, ds\right] dt \\ &= \gamma(D) + \alpha \int_S \int_A q(D|x,a) \\ &\qquad \times \int_0^{\infty} e^{-\alpha t} \int_0^t E_{\gamma}^{\pi}\left[\pi(da|e,s) I_{\{\xi_{s-}(e)\}}(dx)\, ds\right] dt \\ &= \gamma(D) + \frac{1}{\alpha} \int_S \int_A q(D|x,a) \eta^{\pi}(dx,da),\end{aligned}$$

and so (a) follows.

(b) Recall that $\eta(dx,da) = \hat{\eta}(dx)\phi^{\eta}(da|x)$. Then, to prove (b), it suffices to show

$$\int_S \int_A u(x,a)\eta(dx,da) = \int_S \int_A u(x,a)\eta^{\phi^{\eta}}(dx,da) \tag{5.15}$$

for each nonnegative bounded measurable function u on K. In fact, for any such a function u, by Lemma 5.3 in [12] and (2.10) we have

$$\begin{aligned}\alpha V_{\alpha}(x,\phi^{\eta},u) &= \int_{A(x)} u(x,a)\phi^{\eta}(da|x) \\ &\quad + \int_S V_{\alpha}(y,\phi^{\eta},u) q(dy|x,\phi^{\eta}) \qquad \forall x \in S.\end{aligned} \tag{5.16}$$

On the other hand, let $\|u\|_1 := \sup_{(x,a)\in K} |u(x,a)| < \infty$, and $|q(dx|x,\phi^{\eta})|$ the total variation of $q(dy|x,\phi^{\eta})$. Then, by (T_2)–(T_3) and the condition in (b) we have

$$\int_S \int_S |V_{\alpha}(y,\phi^{\eta},u)||q(dy|x,\phi^{\eta})|\hat{\eta}(dx) \le \frac{2\|u\|_1}{\alpha} \int_S |q(\{x\}|x,\phi^{\eta})|\hat{\eta}(dx) < \infty,$$

which together with the Jordan decomposition of $q(\cdot|x,\phi^{\eta})$ and Theorem 2.6.4 in [3], implies

$$\int_S \int_S [\hat{\eta}(dy) q(dx|y,\phi^{\eta})] V_{\alpha}(x,\phi^{\eta},u) = \int_S \left[\int_S V_{\alpha}(y,\phi^{\eta},u) q(dy|x,\phi^{\eta})\right] \hat{\eta}(dx).$$

Hence, by Assumption A(3) we have

$$\begin{aligned}&\lim_{k\to\infty} \int_{S_k} \int_S [\hat{\eta}(dy) q(dx|y,\phi^{\eta})] V_{\alpha}(x,\phi^{\eta},u) \\ &\qquad = \lim_{k\to\infty} \int_{S_k} \left[\int_S V_{\alpha}(y,\phi^{\eta},u) q(dy|x,\phi^{\eta})\right] \hat{\eta}(dx).\end{aligned} \tag{5.17}$$

Thus, for any fixed $k \geq 1$, since $\sup_{x\in S_k} q^*(x) < \infty$, by (5.16) and (a) we have

$$\begin{aligned}
&\int_{S_k}\int_A u(x,a)\eta(dx,da)\\
&\quad= \int_{S_k}\int_{A(x)} u(x,a)[\hat{\eta}(dx)\phi^\eta(da|x)]\\
&\quad= \int_{S_k}\left[\alpha V_\alpha(x,\phi^\eta,u) - \int_S V_\alpha(y,\phi^\eta,u)q(dy|x,\phi^\eta)\right]\hat{\eta}(dx)\\
&\quad= \alpha\int_{S_k} V_\alpha(x,\phi^\eta,u)\gamma(dx) + \int_{S_k} V_\alpha(y,\phi^\eta,u)\left[\int_S \hat{\eta}(dx)q(dy|x,\phi^\eta)\right]\\
&\qquad - \int_{S_k}\left[\int_S V_\alpha(y,\phi^\eta,u)q(dy|x,\phi^\eta)\right]\hat{\eta}(dx)\\
&\quad= \int_{S_k}\int_A u(x,a)\eta^{\phi^\eta}(dx,da) + \int_{S_k}\left[\int_S \hat{\eta}(dy)q(dx|y,\phi^\eta)\right]V_\alpha(x,\phi^\eta,u)\\
&\qquad - \int_{S_k}\left[\int_S V_\alpha(y,\phi^\eta,u)q(dy|x,\phi^\eta)\right]\hat{\eta}(dx),
\end{aligned}$$

which together with (5.17) gives (5.15).

(c) Since $\phi \in \Pi_s$, by (a) and (3.2) we have

$$\begin{aligned}
\alpha\hat{\eta}^\phi(D) &= \alpha\gamma(D) + \int_S q(D|x,\phi)\hat{\eta}^\phi(dx)\\
&= \alpha\gamma(D) + \int_S\int_A q(D|x,a)[\hat{\eta}^\phi(dx)\phi(da|x)]\\
&\qquad\qquad \forall D \in \mathcal{B}(S) \text{ with } \sup_{x\in D} q^*(x) < \infty.
\end{aligned}$$

Moreover, under Assumptions A, B(2) and B(3), by Theorem 3.3 we have

$$\int_S |q(\{x\}|x,\phi)|\hat{\eta}^\phi(dx) \leq L\left[\alpha\int_S w(x)\gamma(dx) + b\right]\Big/[\alpha(\alpha-\rho)] < \infty. \tag{5.18}$$

Thus, by (b) we see that $\hat{\eta}^\phi(dx)\phi(da|x) = \eta^\phi(dx,da)$, and so (c) follows. □

PROOF OF LEMMA 3.8. (a) Since the first part of (a) follows from (3.4), we need to verify the second part of (a). In fact, for each $\mu \in \mathcal{P}(S\times A)$, by (3.5) we have $\int_S \bar{w}(x)\hat{T}'_{\bar{w}}(\mu)(dx) = \frac{1}{\int_S 1/(\bar{w}(x))\hat{\mu}(dx)} < \infty$, and so the second part of (a) follows.

(b) By (3.4) and (3.5) and a straightforward calculation, we see that (b) is true.

(c) and (d). We prove (c) and (d) together. Suppose that $\eta_k \xrightarrow{\bar{w}} \eta_0$. Take any bounded continuous function u on $S\times A$. Then, since $\bar{w}$ is continuous, by $\eta_k \xrightarrow{\bar{w}}$

η_0 we have

$$\lim_{k\to\infty}\int_{S\times A} v(x,a)\bar{w}(x)\eta_k(dx,da)$$
$$=\int_{S\times A} v(x,a)\bar{w}(x)\eta_0(dx,da) \qquad \text{for } v := u, 1,$$

which together with (3.4), imply

$$\lim_{k\to\infty}\int_{S\times A} u(x,a)T_{\bar{w}}(\eta_k)(dx,da)=\int_{S\times A} u(x,a)T_{\bar{w}}(\eta_0)(dx,da), \tag{5.19}$$

and thus, $T_{\bar{w}}(\eta_k)\xrightarrow{1} T_{\bar{w}}(\eta_0)$.

On the other hand, suppose that $\mu_k \xrightarrow{1} \mu_0$, and pick up any continuous function $u(x,a)$ on $S\times A$ such that $|u(x,a)|\le L_u\bar{w}(x)$ for all $(x,a)\in K$, with some nonnegative constant L_u depending on u. Then, the functions $\frac{u(x,a)}{\bar{w}(x)}$ and $\frac{1}{\bar{w}}$ are bounded continuous on $S\times A$. Hence, a straightforward calculation gives

$$\lim_{k\to\infty}\int_{S\times A} u(x,a)T'_{\bar{w}}(\mu_k)(dx,da)=\int_{S\times A} u(x,a)T'_{\bar{w}}(\mu_0)(dx,da). \tag{5.20}$$

By (5.19) and (5.20) and (b), we see that (c) and (d) are both true. □

PROOF OF LEMMA 3.9. (a) For any $\eta^{\pi_1},\eta^{\pi_2}\in\mathcal{M}_o$ and $0\le\beta\le 1$, let $\eta := \beta\eta^{\pi_1}+(1-\beta)\eta^{\pi_2}$. Then, by Theorem 3.5(a) and a straightforward calculation we have

$$\alpha\hat{\eta}(D)=\alpha\gamma(D)+\int_{S\times A} q(D|x,a)\eta(dx,da) \qquad \forall D\in\mathcal{B}(S) \text{ with } \sup_{x\in D} q^*(x)<\infty, \tag{5.21}$$

and also $\int_S w(x)\hat{\eta}(dx)=\int_S w(x)[\beta\hat{\eta}^{\pi_1}(dx)+(1-\beta)\hat{\eta}^{\pi_2}(dx)]<\infty$. Thus, by Theorem 3.5(b) and (5.21), there exists a randomized stationary policy $\phi^\eta\in\Pi_s$ such that $\eta=\eta^{\phi^\eta}$. Hence, $\mathcal{M}_o$ is convex, and thus so is $\mathcal{M}_o^c$.

(b) Take any sequence $\{\eta_m\}$ in $\mathcal{M}_o$ such that $\eta_m \xrightarrow{w} \eta_0$ (and thus $\eta_m \xrightarrow{1} \eta_0$). Then, under Assumptions A, B(2) and B(3), by Theorem 3.1(b) we have

$$\int_S w(x)\hat{\eta}_m(dx)=\int_S w(x)\eta_m(dx,da)\le\frac{\alpha\int_S w(x)\gamma(dx)+b}{\alpha(\alpha-\rho)}$$
$$=M_1^*<\infty \qquad \forall m\ge 1. \tag{5.22}$$

Thus, by Lemma 11.4.7 in [22] we have

$$\int_S |q(\{x\}|x,\phi^{\eta_0})|\hat{\eta}_0(dx)\le L\int_S w(x)\hat{\eta}_0(dx)\le L\liminf_{m\to\infty}\int_S w(x)\hat{\eta}_m(dx)$$
$$\le LM_1^*<\infty.$$

Thus, to prove $\eta_0 \in \mathcal{M}_o$, by Theorem 3.5(b) it suffices to show

$$\alpha\hat{\eta}_0(D) = \alpha\gamma(D) + \int_K q(D|x,a)\eta_0(dx,da)$$
$$\forall D \in \mathcal{B}(S) \text{ with } \sup_{x\in D} q^*(x) < \infty,$$

which can follow (by Proposition 7.18 in [4]) from

$$\alpha \int_S g(y)\hat{\eta}_0(dy) = \alpha \int_S g(y)\gamma(dy) + \int_S \int_K g(y)q(dy|x,a)\eta_0(dx,da) \tag{5.23}$$
$$\forall g \in C_b(S).$$

Thus, the rest verifies (5.23). For any $g \in C_b(S)$, by $\eta_m \in \mathcal{M}_o$ and Theorem 3.5(a) we have

$$\alpha \int_{S_k} g(y)\hat{\eta}_m(dy) = \alpha \int_{S_k} g(y)\gamma(dy) + \int_{S_k} \int_K g(y)q(dy|x,a)\eta_m(dx,da) \tag{5.24}$$
$$\forall k, m \geq 1.$$

Since $q^*(x) \leq Lw(x)$ for all $x \in S$, using Assumption A(3) and the dominated convergence theorem, by (5.22) and (5.24) with letting $k \to \infty$ we have

$$\alpha \int_S g(y)\hat{\eta}_m(dy) = \alpha \int_S g(y)\gamma(dy) + \int_S \int_K g(y)q(dy|x,a)\eta_m(dx,da) \tag{5.25}$$
$$\forall m \geq 1.$$

On the other hand, since $|\int_S g(y)q(dy|x,a)| \leq 2\|g\|_1 q^*(x) \leq 2L\|g\|_1 w(x)$ [for all $a \in A(x)$], by $\eta_m \xrightarrow{w} \eta_0$ and Assumption C(1), we have

$$\lim_{m\to\infty} \int_S g(y)\hat{\eta}_m(dy) = \lim_{m\to\infty} \int_S g(y)\eta_m(dy,da) = \int_S g(y)\eta_0(dy,da)$$
$$= \int_S g(y)\hat{\eta}_0(dy)$$

and

$$\lim_{m\to\infty}\left[\int_S \int_K g(y)q(dy|x,a)\eta_m(dx,da)\right] = \int_S \int_K g(y)q(dy|x,a)\eta_0(dx,da),$$

which together with (5.25) give (5.23), and so (b) follows. □

PROOF OF THEOREM 3.11. (a) Since $\mathcal{P}(S \times A)$ is metrizable, it follows from Lemma 3.8 (with $\bar{w} := w$) that $\mathcal{P}_w(S \times A)$ is also metrizable, and so are $\mathcal{M}_o$ and $\mathcal{M}_o^c$. Since $\mathcal{M}_o$ is closed (by Lemma 3.9) and $\mathcal{M}_o^c$ is a closed subset of $\mathcal{M}_o$ under the additional Assumption C(1), it suffices to show that $\mathcal{M}_o$ is sequentially relatively compact. Indeed, for each $\eta \in \mathcal{M}_o$, since $1 \leq \int_S w'(x)\hat{\eta}(dx) < \infty$ [using

Assumption C(2)], $T_{w'}(\eta)$ is well defined. Moreover, by (3.4) and Theorem 3.3, we have

$$\int_{S\times A}\frac{w(x)}{w'(x)}T_{w'}(\eta)(dx,da)=\frac{\int_{S\times A}w(x)\eta(dx,da)}{\int_{S\times A}w'(x)\eta(dx,da)}$$
$$\le\int_{S\times A}w(x)\eta(dx,da)\le\alpha M_1^*\qquad\forall\eta\in\mathcal{M}_o,$$

where M_1^* is as in Theorem 3.3(b). Thus, by Assumption C(2) and Prohorov' theorem (see Theorem 12.2.15 in [22]) we see that $\{T_{w'}(\eta),\eta\in\mathcal{M}_o\}$ is sequentially relatively compact, and so is $\mathcal{M}_o$ (by Lemma 3.8 with $\bar{w}:=w'$).

(b) Under Assumptions A and B, by Theorem 3.3(b) we have $|V_r(\pi)|\le MM_1^*$ and $|V_n(\pi)|\le MM_1^*$ for $1\le n\le N$. Moreover, by Theorem 3.5 and (2.12) [equivalently, (3.3)] we can find a sequence $\{\eta^{\pi_k}\}$ $(\pi_k\in\Pi_s,k=1,\ldots)$ such that

$$V_r(U)=\lim_{k\to\infty}\frac{1}{\alpha}\int_K r(x,a)\eta^{\pi_k}(dx,da),$$
$$\int_K c_n(x,a)\eta^{\pi_k}(dx,da)\le\alpha d_n,\qquad n=1,\ldots,N.\tag{5.26}$$

Then, by (a) there exists a subsequence $\{\eta^{\pi_{k_m}}\}$ and $\eta_0\in\mathcal{M}_o$ such that $\eta^{\pi_{k_m}}\xrightarrow{w}\eta_0$ as $m\to\infty$, which together with (5.26) implies

$$V_r(U)=\frac{1}{\alpha}\int_K r(x,a)\eta_0(dx,da)$$

and

$$\int_K c_n(x,a)\eta_0(dx,da)\le\alpha d_n,\qquad n=1,\ldots,N,$$

and so ϕ^{η_0} is constrained optimal. □

PROOF OF THEOREM 3.13. Obviously, parts (a), (b) are directive consequence of (3.9) and Theorem 3.5. Moreover, (c) follows from (b) and Theorem 3.11(b). □

PROOF OF THEOREM 3.15. (a) Under Assumptions A, B(2), B(3) and C(3), by Theorems 3.1 and 3.5 we have

$$\mathcal{M}_o=\left\{\eta^\pi\,\Big|\int_S w(x)\hat{\eta}^\pi(dx)\le\alpha M_1^*,\pi\in\Pi\right\}$$
$$=\left\{\eta^\pi\,\Big|\int_S w(x)\hat{\eta}^\pi(dx)\le\alpha M_1^*,\pi\in\Pi_s\right\}.$$

We now prove that η^f is an extreme point in $\mathcal{M}_o$ for each $f\in F$. In fact, for any fixed $f\in F$, suppose that η^f is not any extreme in $\mathcal{M}_o$. Then, there exist $\beta\in(0,1)$ and $\pi_1,\pi_2\in\Pi_s$ such that

$$\eta^f=\beta\eta^{\pi_1}+(1-\beta)\eta^{\pi_2}\quad\text{and}\quad\eta^{\pi_1}\ne\eta^{\pi_2},\tag{5.27}$$

which implies that $\hat{\eta}^{\pi_k} \ll \hat{\eta}^f$ $(k = 1, 2)$. Thus, it follows from (5.27) and Theorem 3.5 that

$$
\begin{aligned}
f(da|x) &= \beta \frac{d\hat{\eta}^{\pi_1}}{d\hat{\eta}^f}(x)\pi_1(da|x) + (1-\beta)\frac{d\hat{\eta}^{\pi_2}}{d\hat{\eta}^f}(x)\pi_2(da|x) \quad \text{and} \\
\beta \frac{d\hat{\eta}^{\pi_1}}{d\hat{\eta}^f}(x) &+ (1-\beta)\frac{d\hat{\eta}^{\pi_2}}{d\hat{\eta}^f}(x) = 1 \qquad \forall x \in \hat{S}
\end{aligned} \tag{5.28}
$$

for some $\hat{S} \in \mathcal{B}(S)$ with $\hat{\eta}^f(\hat{S}) = 1$, where $\frac{d\hat{\eta}^{\pi_k}}{d\hat{\eta}^f}$ denote the (nonnegative) Radon–Nikodym derivative. Moreover, by $\eta^{\pi_1} \neq \eta^{\pi_2}$ we see that $\hat{\eta}^f(\{x \in \hat{S} | \pi_1(\Gamma|x) \neq \pi_2(\Gamma|x)$ for some $\Gamma \in \mathcal{B}(A)\}) > 0$. (Otherwise, η^{π_1} and η^{π_2} coincide.) Thus, for each $x \in \{x \in \hat{S} | \pi_1(\Gamma|x) \neq \pi_2(\Gamma|x)$ for some $\Gamma \in \mathcal{B}(A)\}$, there exists a corresponding $\Gamma_x \in \mathcal{B}(A)$ (depending on x) such that $0 < \pi_1(\Gamma_x|x) < \pi_2(\Gamma_x|x) < 1$. Therefore, by (5.28) we have that $0 < \pi_1(\Gamma_x|x) \leq f(\Gamma_x|x) \leq \pi_2(\Gamma_x|x) < 1$, which contracts with the nonrandom of $f \in F$.

(b) By (a) we only need to show the necessity part. Suppose that $\pi \in \Pi_s$ and $\eta^\pi \neq \eta^f$ for all $f \in F$. Then, there exists $D \in \mathcal{B}(S)$ such that $0 < \hat{\eta}^\pi(D) < 1$ and $0 < \pi(\Gamma_x|x) < 1$ for all $x \in D$ and some $\Gamma_x \in \mathcal{B}(A(x))$ (depending on x). Then, by the condition in (b), there exists $x' \in D$ such that

$$
\begin{aligned}
&0 < \hat{\eta}^\pi(\{x'\}) < 1 \quad \text{and} \\
&0 < \pi(\Gamma_{x'}|x') < 1 \qquad \text{for some } \Gamma_{x'} \in \mathcal{B}(A(x')).
\end{aligned} \tag{5.29}
$$

By (5.29), we now define two policies π_1 and π_2 as follows:

$$
\pi_1(da|x) := \begin{cases} \pi(da|x), & \text{if } x \neq x', \\ \pi(da \cap \Gamma_{x'}|x')/\pi(\Gamma_{x'}|x'), & \text{if } x = x'; \end{cases} \tag{5.30}
$$

$$
\pi_2(da|x) := \begin{cases} \pi(da|x), & \text{if } x \neq x', \\ \pi(da \cap \Gamma_{x'}^c|x')/\pi(\Gamma_{x'}^c|x'), & \text{if } x = x'. \end{cases} \tag{5.31}
$$

Let $\beta := \pi(\Gamma_{x'}|x')$, $\delta' := \frac{\beta\hat{\eta}^{\pi_2}(\{x'\})}{\beta\hat{\eta}^{\pi_2}(\{x'\})+(1-\beta)\hat{\eta}^{\pi_1}(\{x'\})}$ when $\hat{\eta}^{\pi_1}(\{x'\}) + \hat{\eta}^{\pi_1}(\{x'\}) > 0$, and $\delta' = \frac{1}{2}$ when $\hat{\eta}^{\pi_1}(\{x'\}) + \hat{\eta}^{\pi_1}(\{x'\}) = 0$. Then, for each $D \in \mathcal{B}(S)$ with $\sup_{x \in D} q^*(x) < \infty$, by Theorem 3.5 and (5.30), (5.31) as well as a straightforward calculation we have

$$
\begin{aligned}
\alpha\hat{\eta}^{\pi_1}(D) &= \alpha\gamma(D) + \int_{S-\{x'\}} q(D|x,\pi)\hat{\eta}^{\pi_1}(dx) \\
&\quad + \int_{\Gamma_{x'}} q(D|x',a)\pi(da|x')\hat{\eta}^{\pi_1}(\{x'\})/\beta, \\
\alpha\hat{\eta}^{\pi_2}(D) &= \alpha\gamma(D) + \int_{S-\{x'\}} q(D|x,\pi)\hat{\eta}^{\pi_2}(dx) \\
&\quad + \int_{\Gamma_{x'}^c} q(D|x',a)\pi(da|x')\hat{\eta}^{\pi_2}(\{x'\})/(1-\beta).
\end{aligned}
$$

Multiplying by δ' and $(1-\delta')$ the two equalities, respectively, and then summarizing, we have

$$\begin{aligned}&\alpha[\delta'\hat{\eta}^{\pi_1}(D)+(1-\delta')\hat{\eta}^{\pi_2}(D)]\\&\quad=\alpha\gamma(D)+\int_S q(D|x,\pi)[\delta'\hat{\eta}^{\pi_1}(dx)+(1-\delta')\hat{\eta}^{\pi_2}(dx)],\end{aligned}$$

which together with Theorem 3.5(c) implies $\eta^\pi=\delta'\eta^{\pi_1}+(1-\delta')\eta^{\pi_2}$. Moreover, by (5.29) we see that $0<\eta^{\pi_1}(\{x'\}\times\Gamma_{x'})=\hat{\eta}^{\pi_1}(\{x'\})<1$ and $\eta^{\pi_2}(\{x'\}\times\Gamma_{x'})=\hat{\eta}^{\pi_2}(\{x'\})\pi_2(\Gamma_{x'}|x')=0$. Hence, $\eta^\pi=\delta'\eta^{\pi_1}+(1-\delta')\eta^{\pi_2}$ is not an extreme point. □

PROOF OF THEOREM 3.16. Let ϕ^* be a constrained optimal policy [by Theorem 3.13(c)], and $\mathcal{M}_o^c(e)$ be the set of all extreme points in $\mathcal{M}_o^c$ in (3.7). Since $\mathcal{M}_o^c$ has been proved to be convex compact [by Theorem 3.11(a) and Lemma 3.9]. Thus, by Choquet's theorem [32], η^{ϕ^*} is the barycenter of a probability measure $\bar{\mu}$ supported on $\mathcal{M}_o^c(e)$. Therefore,

$$\int_{S\times A}c_0(x,a)\eta^{\phi^*}(dx,da)=\int_{\mathcal{M}_o^c(e)}\left(\int_{S\times A}c_0(x,a)\eta(dx,da)\right)\bar{\mu}(d\eta). \tag{5.32}$$

On the other hand, since $\int_{S\times A}c_0(x,a)\eta^{\phi^*}(dx,da)\le\int_{S\times A}c_0(x,a)\eta(dx,da)$ for all $\eta\in\mathcal{M}_o^c(e)$, it follows from (5.32) that there exists $\eta^*\in\mathcal{M}_o^c(e)$ such that

$$\int_{S\times A}c_0(x,a)\eta^{\phi^*}(dx,da)=\int_{S\times A}c_0(x,a)\eta^*(dx,da).$$

Hence, $\pi^*:=\phi^{\eta^*}$ is also constrained optimal. Moreover, since $\int_{S\times A}c_n(x,a)\eta(dx,da)$ (for each fixed $1\le n\le N$) is linear in $\eta\in\mathcal{M}_o$ and thus can be regarded as a "hyperplane," each extreme point of $\mathcal{M}_o^c$ is a convex combination of at most $N+1$ extreme points in $\mathcal{M}_0$. That is, there exists $(N+1)$ numbers $p_k\ge0$ and stationary policies $f_k\in F$ $(k=1,\dots,N+1)$ (using Theorem 3.15) such that $\eta^*=p_1\eta^{f_1}+\cdots+p_{N+1}\eta^{f_{N+1}}$, $p_1+\cdots+p_{N+1}=1$, which together with Theorem 3.15 and (3.2) completes the proof of this theorem. □

REFERENCES

[1] ALTMAN, E. (1999). *Constrained Markov Decision Processes*. Chapman and Hall/CRC, Boca Raton, FL. MR1703380

[2] ALTMAN, E. and SHWARTZ, A. (1991). Markov decision problems and state-action frequencies. *SIAM J. Control Optim.* **29** 786–809. MR1111660

[3] ASH, R. B. (2000). *Probability and Measure Theory*, 2nd ed. Academic Press, Burlington, MA. MR1810041

[4] BERTSEKAS, D. P. and SHREVE, A. (1996). *Stochastic Optimal Control*: *The Case of Discrete-Time Case*. Athena Scientific, Belmont, MA.

[5] CHEN, M.-F. (2004). *From Markov Chains to Non-equilibrium Particle Systems*, 2nd ed. World Scientific, River Edge, NJ. MR2091955

[6] CHEN, R. C. and FEINBERG, E. A. (2007). Non-randomized policies for constrained Markov decision processes. *Math. Methods Oper. Res.* **66** 165–179. MR2317865
[7] FEINBERG, E. A. (2000). Constrained discounted Markov decision processes and Hamiltonian cycles. *Math. Oper. Res.* **25** 130–140. MR1854324
[8] FEINBERG, E. A. and SHWARTZ, A. (1995). Constrained Markov decision models with weighted discounted rewards. *Math. Oper. Res.* **20** 302–320. MR1342949
[9] FEINBERG, E. A. and SHWARTZ, A. (1996). Constrained discounted dynamic programming. *Math. Oper. Res.* **21** 922–945. MR1419909
[10] FEINBERG, E. A. and SHWARTZ, A. (1999). Constrained dynamic programming with two discount factors: Applications and an algorithm. *IEEE Trans. Automat. Control* **44** 628–631. MR1680195
[11] GUO, X. P. (2007). Constrained optimization for average cost continuous-time Markov decision processes. *IEEE Trans. Automat. Control* **52** 1139–1143. MR2329912
[12] GUO, X. P. (2007). Continuous-time Markov decision processes with discounted rewards: The case of Polish spaces. *Math. Oper. Res.* **32** 73–87. MR2292498
[13] GUO, X. P. and HERNÁNDEZ-LERMA, O. (2003). Constrained continuous-time Markov control processes with discounted criteria. *Stochastic Anal. Appl.* **21** 379–399. MR1967719
[14] GUO, X. P. and HERNÁNDEZ-LERMA, O. (2003). Continuous-time controlled Markov chains. *Ann. Appl. Probab.* **13** 363–388. MR1952002
[15] GUO, X. P. and HERNÁNDEZ-LERMA, O. (2009). *Continuous-time Markov Decision Processes*: *Theory and Applications*. *Stochastic Modelling and Applied Probability* **62**. Springer, Berlin. MR2554588
[16] GUO, X. P. and RIEDER, U. (2006). Average optimality for continuous-time Markov decision processes in Polish spaces. *Ann. Appl. Probab.* **16** 730–756. MR2244431
[17] GUO, X. P., HERNÁNDEZ-LERMA, O. and PRIETO-RUMEAU, T. (2006). A survey of resent results on continuous-time Markov decision processes. *TOP* **14** 177–246.
[18] HAVIV, M. and PUTERMAN, M. L. (1998). Bias optimality in controlled queueing systems. *J. Appl. Probab.* **35** 136–150. MR1622452
[19] HERNÁNDEZ-LERMA, O. and GONZÁLEZ-HERNÁNDEZ, J. (2000). Constrained Markov control processes in Borel spaces: The discounted case. *Math. Methods Oper. Res.* **52** 271–285. MR1797253
[20] HERNÁNDEZ-LERMA, O., GONZÁLEZ-HERNÁNDEZ, J. and LÓPEZ-MARTÍNEZ, R. R. (2003). Constrained average cost Markov control processes in Borel spaces. *SIAM J. Control Optim.* **42** 442–468 (electronic). MR1982278
[21] HERNÁNDEZ-LERMA, O. and LASSERRE, J. B. (1996). *Discrete-time Markov Control Processes*: *Basic Optimality Criteria*. *Applications of Mathematics* (*New York*) **30**. Springer, New York. MR1363487
[22] HERNÁNDEZ-LERMA, O. and LASSERRE, J. B. (1999). *Further Topics on Discrete-time Markov Control Processes*. *Applications of Mathematics* (*New York*) **42**. Springer, New York. MR1697198
[23] HORDIJK, A. and SPIEKSMA, F. (1989). Constrained admission control to a queueing system. *Adv. in Appl. Probab.* **21** 409–431. MR0997731
[24] JACOD, J. (1975). Multivariate point processes: Predictable projection, Radon–Nikodým derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* **31** 235–253. MR0380978
[25] KADOTA, Y., KURANO, M. and YASUDA, M. (2006). Discounted Markov decision processes with utility constraints. *Comput. Math. Appl.* **51** 279–284. MR2203079
[26] KAKUMANU, P. (1971). Continuously discounted Markov decision model with countable state and action space. *Ann. Math. Statist.* **42** 919–926. MR0282651
[27] KITAEV, M. Y. (1985). Semi-Markov and Jump Markov controlled models: Average cost criterion. *Theory Probab. Appl.* **30** 272–288.

[28] KITAEV, M. Y. and RYKOV, V. V. (1995). *Controlled Queueing Systems*. CRC Press, Boca Raton, FL. MR1413045

[29] KURANO, M., NAKAGAMI, J.-I. and HUANG, Y. (2000). Constrained Markov decision processes with compact state and action spaces: The average case. *Optimization* **48** 255–269. MR1785314

[30] LEWIS, M. E. and PUTERMAN, M. L. (2001). A probabilistic analysis of bias optimality in unichain Markov decision processes. *IEEE Trans. Automat. Control* **46** 96–100. MR1809468

[31] LUND, R. B., MEYN, S. P. and TWEEDIE, R. L. (1996). Computable exponential convergence rates for stochastically ordered Markov processes. *Ann. Appl. Probab.* **6** 218–237. MR1389838

[32] PHELPS, R. R. (2001). *Lectures on Choquet's Theorem*, 2nd ed. *Lecture Notes in Math.* **1757**. Springer, Berlin. MR1835574

[33] PIUNOVSKIY, A. B. (1997). *Optimal Control of Random Sequences in Problems with Constraints*. *Mathematics and Its Applications* **410**. Kluwer, Dordrecht. MR1472738

[34] PIUNOVSKIY, A. B. (1998). A controlled jump discounted model with constraints. *Theory Probab. Appl.* **42** 51–72.

[35] PIUNOVSKIY, A. B. (2005). *Discounted Continuous Time Markov Decision Processes*: *The Convex Analytic Approach*. 16th Triennial IFAC World Congress, Czech Republic, Praha.

[36] PRIETO-RUMEAU, T. and HERNÁNDEZ-LERMA, O. (2008). Ergodic control of continuous-time Markov chains with pathwise constraints. *SIAM J. Control Optim.* **47** 1888–1908. MR2421334

[37] PUTERMAN, M. L. (1994). *Markov Decision Processes*: *Discrete Stochastic Dynamic Programming*. Wiley, New York. MR1270015

[38] SENNOTT, L. I. (1991). Constrained discounted Markov decision chains. *Probab. Engrg. Inform. Sci.* **5** 463–475. MR1183189

[39] SENNOTT, L. I. (1999). *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley, New York. MR1645435

[40] YUSHKEVICH, A. A. (1977). Controlled Markov models with contable states and continous time. *Theory Probab. Appl.* **22** 215–7235.

[41] ZADOROJNIY, A. and SHWARTZ, A. (2006). Robustness of policies in constrained Markov decision processes. *IEEE Trans. Automat. Control* **51** 635–638. MR2228025

[42] ZHANG, L. L. and GUO, X. P. (2008). Constrained continuous-time Markov decision processes with average criteria. *Math. Methods Oper. Res.* **67** 323–340. MR2390062

SCHOOL OF MATHEMATICS
AND COMPUTATIONAL SCIENCE
ZHONGSHAN UNIVERSITY
GUANGZHOU 510275
P. R. CHINA
E-MAIL: mcsgxp@mail.sysu.edu.cn

DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
LSB114, SHATIN, HONG KONG
P. R. CHINA
E-MAIL: xysong@sta.cuhk.edu.hk