Adv. Appl. Prob. **45,** 490–519 (2013)
Printed in Northern Ireland

ABSORBING CONTINUOUS-TIME MARKOV DECISION PROCESSES WITH TOTAL COST CRITERIA

XIANPING GUO,* *Sun Yat-Sen University*
MANTAS VYKERTAS,** *Open University*
YI ZHANG,*** *University of Liverpool*

Abstract

In this paper we study absorbing continuous-time Markov decision processes in Polish state spaces with unbounded transition and cost rates, and history-dependent policies. The performance measure is the expected total undiscounted costs. For the unconstrained problem, we show the existence of a deterministic stationary optimal policy, whereas, for the constrained problems with N constraints, we show the existence of a mixed stationary optimal policy, where the mixture is over no more than $N + 1$ deterministic stationary policies. Furthermore, the strong duality result is obtained for the associated linear programs.

Keywords: CTMDP; total cost; constrained optimality; linear program

2010 Mathematics Subject Classification: Primary 90C40
Secondary 60J25; 60J75

1. Introduction

Continuous-time Markov decision processes (CTMDPs) have found rich applications to telecommunication, queueing systems, epidemiology, etc.; see the examples in the monographs [13] and [25]. Two standard performance measures of a CTMDP are the (expected) long-run average costs [12], [14], [17], [24], [34], [39] and the (expected) total discounted costs [15], [27], [30]–[32]. The long-run average criteria are not appropriate for CTMDPs with transient behavior because in that case the long-run average costs will be zero for each policy. For short-term decision making, discounted criteria are often employed. For a discounted CTMDP, the (positive) constant discount factor is often understood as the risk-free rate of return at which the interest is continuously compounded. Alternatively, one can regard it as the constant intensity at which the CTMDP jumps (independently of anything else) to an artificially defined absorbing state, where no further cost is incurred and no further transition takes place. Adopting the latter interpretation of the discount factor, the expected total discounted cost is equivalently realized as the expected total undiscounted cost up to the absorbing time; see Section 2 of [9]. In other words, discounted CTMDPs can be recovered from the more general CTMDPs with an absorbing set and total (undiscounted) cost criteria. The opposite direction does not hold.

Received 17 May 2012; revision received 29 August 2012.
* Postal address: School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, P. R. China. Email address: mcsgxp@mail.sysu.edu.cn
** Postal address: Department of Mathematics and Statistics, Open University, Milton Keynes, MK7 6AA, UK. Email address: mantas.vykertas@gmail.com
*** Postal address: Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK. Email address: yi.zhang@liv.ac.uk

Leaving alone their relationship with the discounted CTMDPs, absorbing CTMDPs are also of special interest for their applications too, for instance, epidemic models and population dynamics, where the state (the number of infected population) zero, indicating that the epidemic vanishes, is often taken as the absorbing state, and one is interested in minimizing the expected total undiscounted cost (from immunization, isolation, etc.) up to the absorbing time. For instance, this criterion has been employed in [7] for susceptible-infective-removed (SIR) models (in a deterministic setup) initially considered in [11], and in [29] for a controlled birth-and-death process.

Motivated by the above discussion, in this paper we study CTMDPs in *general (Polish)* state spaces with a measurable absorbing set, *unbounded* transition rates and unbounded (from both above and below) cost rates, *history-dependent policies*, constraints, and total *undiscounted* cost criteria. The main contributions of this paper are as follows. Firstly, for unconstrained CTMDPs, we prove the existence of a deterministic stationary optimal policy, and that the value function is given by the unique solution to the Bellman equation. Secondly, for constrained CTMDPs with an arbitrary number of constraints N, we develop its convex analytic approach. In greater detail, we reformulate the original CTMDPs as convex programs in the space of occupation measures, whose compactness is shown in an appropriate topology, leading to the existence of a randomized stationary (constrained) optimal policy, whose occupation measure, under extra conditions, is then shown to be a convex combination of no more than $N+1$ occupation measures of deterministic stationary policies. Thirdly, we further formulate the CTMDP as an infinite-dimensional primal linear program, and prove its strong duality with the dual linear program.

CTMDPs with total undiscounted cost criteria are considered in [33], requiring uniformly bounded transition rates; in that case, the uniformization technique could be employed. There have been very few articles on CTMDPs with unbounded transition rates and total undiscounted cost criteria, see, for example, [28], which allows one to change instantaneously the state of the controlled process, but excludes the gradual control of the process through the transition rates. In the current literature on CTMDPs with gradual controls, unbounded transition rates, and total undiscounted cost criteria, to the best of the authors' knowledge, there is only one article [16]. However, there are significant differences between [16] and the present article. (i) We consider CTMDPs in Polish spaces and with history-dependent policies, while Guo and Zhang [16] considered the case of countable state spaces and was restricted to the class of randomized Markov policies only. (ii) For the constrained CTMDPs, we allow an arbitrary number of constraints, whereas Guo and Zhang only considered the case of one constraint. Note that our approach is based on the studies of occupation measures, and is different from the approach employed in [16] based on the Lagrange multiplier method, which is not suitable for the case of multiple constraints and Polish state spaces. (iii) We consider the linear programming formulation of the CTMDPs and derive the (strong) duality result, which was not touched on in [16] at all. Therefore, this paper is a significant nontrivial extension of and improvement over [16], and fills the gaps in the current literature about CTMDPs.

Absorbing models in the discrete-time framework have received significant attention in the literature; however, as our interest lies in their continuous-time counterpart, we only refer the interested reader to the monographs [2] and [19], where [2] is in the framework of denumerable state spaces, and [19] is about unconstrained problems only. A very recent contribution to this topic is [10].

Note that, since we allow transition rates to be unbounded, the standard uniformization technique might not be applicable to reducing the CTMDPs to equivalent discrete-time Markov decision processes (DTMDPs).

The rest of this paper is organized as follows. We describe the mathematical model and introduce the terminology in Section 2. We then study the dynamic programming approach for the unconstrained CTMDP problem in Section 3. In Section 4 we present the convex analytic approach for the constrained CTMDP problem. Section 5 includes the linear programming formulation and duality results. We point out that the results obtained in Section 2 for unconstrained CTMDPs, while being interesting in their own right, are also needed for the studies of constrained absorbing CTMDPs in Sections 4 and 5. Examples illustrating possible applications of the obtained results are given in Section 6. Concluding remarks are presented in Section 7.

2. Description of the mathematical model

Notation. Throughout the paper, we only consider finite (signed) measures, and the measurability is always understood in the Borel sense. We denote by $\mathbf{1}_D$ the indicator function of any set D, by $\delta_x(\cdot)$ the Dirac measure concentrated at a point x, by $\mathcal{B}(X)$ the Borel σ-algebra of a topological space X, and by $\bigvee_{0\le t<s}\mathcal{F}_t$ the smallest σ-algebra containing all the σ-algebras $\{\mathcal{F}_t,\ 0\le t<s\}$. We define $\mathbb{R}_+ := (0,\infty)$, $\mathbb{R}_+^0 := [0,\infty)$, and $\mathbb{Z}_+^0 := \{0,1,\dots\}$. The abbreviations 's.t.' and 'a.s.' stand for 'subject to' and 'almost surely'.

2.1. Kitaev's construction of CTMDPs

The primitives of a CTMDP are the elements [15], [24], [25], [30]

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

where

- S (state space) is a nonempty Polish space endowed with the Borel σ-algebra $\mathcal{B}(S)$;
- A (action space) is a nonempty Borel space endowed with the Borel σ-algebra $\mathcal{B}(A)$;
- $A(x)$ (admissible action sets) are nonempty subsets in $\mathcal{B}(A)$ such that the space of admissible state-action pairs $\mathbb{K} := \{(x,a)\in S\times A : a\in A(x)\}$ is a subset in $\mathcal{B}(S\times A)$ and contains the graph of a (Borel) measurable mapping ϕ from S to A such that $\phi(x)\in A(x)$ for all $x\in S$ (to ensure the existence of a deterministic stationary policy);
- $q(\mathrm{d}y\mid x,a)$ (transition rates) is a signed kernel on $\mathcal{B}(S)$ given $(x,a)\in\mathbb{K}$, taking nonnegative values on $\Gamma_S\setminus\{x\}$ for all $\Gamma_S\in\mathcal{B}(S)$, being conservative in the sense of $q(S\mid x,a)=0$, and stable in that $\bar{q}_x=\sup_{a\in A(x)} q_x(a)<\infty$, where $q_x(a) := -q(\{x\}\mid x,a)$;
- $\gamma(\cdot)$ (initial distribution) is a probability measure on $(S,\mathcal{B}(S))$.

Given the aforementioned primitives, one can refer to Kitaev's approach, see [24], for the construction of the underlying stochastic basis $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\ge0},\mathrm{P}_\gamma^\pi)$ and the controlled process $\{\xi_t,\ t\ge0\}$ thereon. Below we briefly recall it in order to define the necessary terminologies; see [15], [24], [25], [30] for more details.

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

$$(x_0,\theta_1,x_1,\dots,\theta_{m-1},x_{m-1},\infty,x_\infty,\infty,x_\infty,\dots),$$

where $x_\infty\notin S$ is an isolated point, $x_l\in S$, $\theta_{l+1}\in\mathbb{R}_+$, $0\le l\le m-1$, and $m\ge1$, we obtain the sample space $(\Omega,\mathcal{F})$, where $\mathcal{F}$ is the standard Borel σ-algebra. For each $m\ge1$, define on Ω the maps $T_0(\omega) := 0$, $T_m(\omega) := \theta_1 + \theta_2 + \cdots + \theta_m$, $T_\infty(\omega) := \lim_{m\to\infty} T_m(\omega)$, $X_m(\omega) := x_m$, and the process of interest $\{\xi_t,\ t \geq 0\}$ by

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

for all $\omega = (x_0, \theta_1, x_1, \ldots, \theta_m, x_m, \ldots) \in \Omega$, where x_∞ is the isolated point and will be regarded as a cemetery since we do not intend to consider the process after T_∞.

To continue describing the construction of CTMDPs, we need to introduce some notation. Let $\mathcal{F}_t := \sigma(\{T_m \leq s, X_m \in \Gamma_S\} \colon \Gamma_S \in \mathcal{B}(S),\ s \leq t,\ m \geq 0)$ for all $t \geq 0$, $A_\infty := A \cup \{a_\infty\}$, $S_\infty := S \cup \{x_\infty\}$, $A(x_\infty) := \{a_\infty\}$, $q_{x_\infty}(a_\infty) = 0$, and $\mathcal{F}_{s-} := \bigvee_{0\leq t<s} \mathcal{F}_t$. The predictable (with respect to $\{\mathcal{F}_t\}_{t\geq 0}$) σ-algebra $\mathbb{P}$ on $\Omega \times \mathbb{R}_+^0$ is given by $\mathbb{P} := \sigma(\Gamma \times \{0\}(\Gamma \in \mathcal{F}_0), \Gamma \times (s, \infty)\ (\Gamma \in \mathcal{F}_{s-}))$; see [25, Chapter 4] for more details. Now the following definitions are in position.

- Randomized history-dependent policy: $\pi(\cdot \mid \omega, t)$, a $\mathbb{P}$-measurable transition probability function on $(A_\infty, \mathcal{B}(A_\infty))$, concentrated on $A(\xi_{t-}(\omega))$.
- Randomized Markov policy: $\pi(\cdot \mid \omega, t) = \pi^M(\cdot \mid \xi_{t-}(\omega), t)$. Here $\pi^M(\cdot \mid x, t)$ is a kernel on A_∞ given $S_\infty \times \mathbb{R}_+^0$.
- (Ordinary) randomized stationary policy: $\pi(\cdot \mid \omega, t) = \pi^S(\cdot \mid \xi_{t-}(\omega))$. Here $\pi^S(\cdot \mid x)$ is a kernel on A_∞ given S_∞.
- Deterministic stationary policy: $\pi(\cdot \mid \omega, t) = \mathbf{1}_{\{\cdot\}}(\phi(\xi_{t-}(\omega)))$, where $\phi \colon S_\infty \to A_\infty$ is a measurable mapping such that $\phi(x) \in A(x)$ for all $x \in S_\infty$. Such policies are denoted as ϕ.

Below we denote by Π_{H} the class of randomized history-dependent policies, and by Π_{S} the class of randomized stationary policies.

Under any fixed policy $\pi \in \Pi_{\mathrm{H}}$, let us define

$$\nu^\pi(\omega, \mathrm{d}t \times \Gamma_S) := \left[\int_A \pi(\mathrm{d}a \mid \omega, t) q(\Gamma_S \setminus \{\xi_{t\_}(\omega)\} \mid \xi_{t-}(\omega), a)\right] \mathrm{d}t$$

for any $\Gamma_S \in \mathcal{B}(S)$. This random measure is predictable, and such that $\nu^\pi(\omega, \{t\} \times S) = \nu^\pi(\omega, [T_\infty, \infty) \times S) = 0$; see [24], [25], and [27]. Therefore, by [24], see also [23] and [25, Chapter 4], there exists a unique probability measure P_γ^π on $(\Omega, \mathcal{F})$ such that $\mathrm{P}_\gamma^\pi(\xi_0 \in \mathrm{d}x) = \gamma(\mathrm{d}x)$, and, with respect to P_γ^π, ν^π is the dual predictable projection of the random measure $\mu(\mathrm{d}t, \mathrm{d}y) := \sum_{m\geq 1} \mathbf{1}_{\{T_m<\infty\}}\, \mathbf{1}_{\mathrm{d}y}(X_m)\, \mathbf{1}_{\mathrm{d}t}(T_m)$. This gives rise to the desired stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathrm{P}_\gamma^\pi)$, always assumed to be complete, concluding Kitaev's construction.

Below, when $\gamma(\cdot)$ is a Dirac measure concentrated at $x \in S$, we use the 'degenerated' notation P_x^π. Expectations with respect to P_γ^π and P_x^π are denoted as E_γ^π and E_x^π, respectively.

2.2. Absorbing CTMDP models

Absorbing set and further notation. In order to state the CTMDP optimization problem under consideration, we consider measurable functions $c_i(x, a)$, $i = 0, 1, \ldots, N$, on $\mathbb{K}$, representing the cost rates, and fixed constants d_j, $j = 1, \ldots, N$. In this paper we are particularly interested in CTMDPs with an absorbing set, namely, $\Delta \in \mathcal{B}(S)$ such that, for each $x \in \Delta$, $A(x) := A$, $c_i(x, a) = 0$, $i = 0, 1, \ldots, N$, for all $a \in A(x)$, and $q(\Gamma_S \mid x, a) = 0$ for all $\Gamma_S \in \mathcal{B}(S)$ and $a \in A(x)$. In other words, once the CTMDP enters the set Δ, it remains there, and no

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

further cost will be incurred thereafter. In what follows, we use the notation $X := S \setminus \Delta$ and $K := \mathbb{K} \setminus (\Delta \times A)$ for brevity. Since Δ and $\mathbb{K}$ are measurable by assumption, so are the sets X and K. Throughout this paper, by an absorbing CTMDP model we mean the collection $\{S, A, A(x), q(\mathrm{d}y \mid x, a), c_0, (c_j, d_j)_{j=1}^k, \gamma, \Delta\}$.

As is well known, in general, we may have $\mathrm{P}_\gamma^\pi(T_\infty < \infty) > 0$, which implies that the process $\{\xi_t,\ t \geq 0\}$ is explosive. To avoid such explosiveness, we impose the following condition.

Condition 1. *There exist constants $\rho > 0$ and $L \geq 0$, and a measurable function w on S satisfying $w(x) \geq 1$ for each $x \in X$ and $w(x) = 0$ for each $x \in \Delta$ such that*

(a) $\int_S q(\mathrm{d}y \mid x, a) w(y) \leq -\rho w(x)$ *for all $x \in X, a \in A(x)$;*

(b) $\overline{q}_x \leq L w(x)$ *for each $x \in X$.*

Note that, under Condition 1, for the increasing system of measurable subsets $X_l \subseteq X$ defined by $X_l := \{x \in X : w(x) \leq l\}$, $l = 1, 2, \ldots$, it holds that $\bigcup_{l=0}^\infty X_l = X$ and $\lim_{l\to\infty} \inf_{x\in X\setminus X_l} w(x) = \infty$.

Condition 1 guarantees that, under any policy π and given any initial state $x \in S$, the expected total time before the controlled process gets absorbed is finite; see the discussion immediately after Definition 2 and (7) below. On the other hand, the examples presented in Section 6 illustrate that Condition 1 also admits CTMDPs with unbounded transition rates. The case of bounded transition rates is of less interest as then one can apply the uniformization technique to pass to the equivalent DTMDPs.

Under Condition 1, we have the following lemma.

Lemma 1. *Suppose that Condition 1 is satisfied. Then the following assertions hold.*

(a) *For each $\pi \in \Pi_{\mathrm{H}}$, $\mathrm{P}_\gamma^\pi(T_\infty = \infty) = 1$, and then $\mathrm{P}_\gamma^\pi(\xi_t \in S) = 1$ for all $t \geq 0$.*

(b) $\mathrm{E}_x^\pi[w(\xi_t)] \leq \mathrm{e}^{-\rho t} w(x)$ *for all $\pi \in \Pi_{\mathrm{H}}$, $x \in S$, and $t \geq 0$.*

Proof. The statements follow from Theorem 1(a) of [30], where the requirement of $w(x) \geq 1$ for each $x \in S$ can be replaced with that of $w \geq 0$ without violating its proof.

According to Lemma 1(a), the explosion does not happen to the CTMDP under every history-dependent policy π. Throughout this paper, we always assume that Condition 1 is satisfied.

In order for the CTMDP optimization problem (yet to be introduced) to be well defined, we impose the next condition.

Condition 2. (a) $\int_S w(x)\gamma(\mathrm{d}x) < \infty$, *where the function w comes from Condition 1.*

(b) *There exists a constant $M \geq 0$ such that $|c_i(x, a)| \leq M w(x)$ for each $x \in S$, $a \in A(x)$, and $i = 0, 1, \ldots, N$.*

Under Condition 2, we have the following statement.

Lemma 2. *Suppose that Conditions 1 and 2(b) are satisfied. Then the following assertions hold.*

(a) *For each $x \in S$, $\pi \in \Pi_{\mathrm{H}}$, and $i = 0, 1, \ldots, N$,*

$$\mathrm{E}_x^\pi\left[\int_0^\infty \int_A |c_i(\xi_{t-}, a)| \pi(\mathrm{d}a \mid \omega, t)\, \mathrm{d}t\right] \leq \frac{M}{\rho} w(x).$$

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

(b) *If, additionally, Condition 2(a) is also satisfied then*

$$\mathrm{E}_\gamma^\pi\left[\int_0^\infty\int_A |c_i(\xi_{t-},a)|\pi(\mathrm{d}a\mid\omega,t)\,\mathrm{d}t\right]<\infty.$$

Proof. The proof is an immediate consequence of Lemma 1(b).

Remark 1. In what follows, for simplicity, we put ξ_t in place of ξ_{t-} in formulae like those in Lemma 2; obviously, this does not change the values of the underlying functionals.

Below, for any given measurable function g on $\mathbb{K}$, we use the notation

$$V(\gamma,\pi,g):=\mathrm{E}_\gamma^\pi\left[\int_0^\infty\int_A g(\xi_t,a)\pi(\mathrm{d}a\mid\omega,t)\,\mathrm{d}t\right]$$

whenever the right-hand side is well defined. When $\gamma(\mathrm{d}y)=\delta_x(\mathrm{d}y)$, where $\delta_x(\mathrm{d}y)$ is the Dirac measure concentrated on $\{x\}$, we use the simpler notation $V(x,\pi,g)$ instead of $V(\delta_x,\pi,g)$.

Under Conditions 1 and 2, the following absorbing CTMDP optimization problems of our interest are well defined.

Unconstrained absorbing CTMDP problem:

$$V(x,\pi,c_0)\to\min_{\pi\in\Pi_\mathrm{H}}\qquad\text{for all } x\in S. \tag{1}$$

Constrained absorbing CTMDP problem:

$$\begin{aligned} &V(\gamma,\pi,c_0)\to\min_{\pi\in\Pi_\mathrm{H}}\\ \text{s.t.}\quad &V(\gamma,\pi,c_j)\le d_j,\qquad j=1,2,\dots,N. \end{aligned}\tag{2}$$

In order to discuss the solvability of the above absorbing CTMDP optimization problems, we state the following definition.

Definition 1. (a) For the unconstrained CTMDP problem (1), a policy $\pi^*\in\Pi_\mathrm{H}$ is said to be optimal if $V(x,\pi^*,c_0)=\inf_{\pi\in\Pi_\mathrm{H}}V(x,\pi,c_0)$ for each $x\in S$.

(b) For the constrained CTMDP problem (2), a policy $\pi^*\in\Pi_\mathrm{H}$ is said to be feasible if $V(\gamma,\pi^*,c_j)\le d_j$ for each $j=1,\dots,N$. Denoting by Π_F the set of feasible policies, a feasible policy π^* is said to be (constrained) optimal if $V(\gamma,\pi,c_0)=\inf_{\pi\in\Pi_\mathrm{F}}V(\gamma,\pi,c_0)$.

The main goal here is to give the existence of optimal policies and the linear programming formulations for the unconstrained and constrained CTMDP problems above.

3. Dynamic programming for unconstrained CTMDPs

In this section we show that the value function of problem (1) can be obtained by solving the Bellman equation, and there exists a deterministic stationary optimal policy out of the class of (randomized) history-dependent ones. To this end, we need to impose some further conditions.

Condition 3. *There exist constants $L'\ge 0$, $\rho'>0$, and $M'\ge 0$, and a measurable function w' on S satisfying $w'(x)\ge 1$ for each $x\in X$ and $w'(x)=0$ for each $x\in\Delta$ such that*

(a) $(\overline{q}_x+1)w'(x)\le L'w(x),\ x\in X;$

(b) $\int_S q(\mathrm{d}y\mid x,a)w'(y)\le-\rho'w'(x),\ x\in X,$ *and* $a\in A(x)$*;*

(c) $|c_i(x,a)|\le M'w'(x)$ *for each* $i=0,1,\dots,N$, $x\in X$, *and* $a\in A(x)$.

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

Condition 3 (particularly parts (a) and (b)) validates Dynkin's formula stated in Lemma 3 below, which is used in the proof of Theorem 1 below. Condition 3(c) admits possibly unbounded (from both above and below) cost rates since the function w' could be unbounded on S. Moreover, under Conditions 1 and 3, it can be shown that part (a) of Lemma 2 still holds with w, ρ, and M replaced by w', ρ', and M'. Finally, when Condition 3 is satisfied, Condition 2(b) automatically follows, and thus can be omitted.

Lemma 3. *Suppose that Conditions 1 and 3(a), (b) are satisfied. Then the following Dynkin's formula holds for any $x \in S$, $\pi \in \Pi_{\mathrm{H}}$, $t \geq 0$, and w'-bounded function u on S (i.e. the measurable function u satisfies $\sup_{x\in X} |u(x)|/w'(x) < \infty$ and $u(x) = 0$ for each $x \in \Delta$):*

$$\mathrm{E}_x^{\pi}[u(\xi_t)] - u(x) = \mathrm{E}_x^{\pi}\left[\int_0^t \int_S \int_A \pi(\mathrm{d}a \mid \omega, v) q(\mathrm{d}y \mid \xi_v, a) u(y)\, \mathrm{d}v\right].$$

Proof. We define $\tilde{w}(x) := w(x) + \mathbf{1}_{\Delta}(x)$, $\tilde{w}'(x) := w'(x) + \mathbf{1}_{\Delta}(x)$, and $\tilde{S}_l := X_l \cup \Delta$, where X_l is as in the discussion immediately after Condition 1. Then, under the conditions of the statement, it is easy to verify that Conditions 1 and 5(a), (b) of [30] are satisfied by $\tilde{w}$, $\tilde{S}_l$, and $\tilde{w}'$. Hence, the statement follows from Theorem 3 of [30].

Condition 4. (a) *For any bounded measurable function u on X and for fixed $x \in X$, $\int_X u(y)q(\mathrm{d}y \mid x, a)$ is lower semicontinuous in $a \in A(x)$.*

(b) *For each fixed $x \in X$, $\int_X w(y)q(\mathrm{d}y \mid x, a)$ is continuous in $a \in A(x)$.*

(c) *For each $i = 0, 1, \ldots, N$ and $x \in X$, $c_i(x, a)$ is lower semicontinuous in $a \in A(x)$.*

(d) *For each fixed $x \in X$, the set $A(x)$ is compact in A. (Recall that $X = S \setminus \Delta$.)*

Condition 4 is a standard compactness-continuity condition, and a counterpart of Assumptions 8.3.1 and 8.3.3 of [19] imposed for DTMDPs. Condition 4(a) is obviously equivalent to saying that $\int_X u(y)q(\mathrm{d}y \mid x, a)$ is continuous in $a \in A(x)$ for each $x \in S$ and bounded measurable function u on X. Moreover, Condition 4(a) and (b) are equivalent to saying that, for each $x \in X$, $\int_X u(y)q(\mathrm{d}y \mid x, a)$ is continuous in $a \in A(x)$ for each w-bounded function u on S (i.e. the measurable function u satisfies $\sup_{x\in X} |u(x)|/w(x) < \infty$ and $u(x) = 0$ for each $x \in \Delta$). Indeed, if we suppose that Condition 4(a) and (b) are satisfied, and consider an arbitrarily fixed w-bounded function u on S and the probability measure $q(\mathrm{d}y \mid x, a)/(1 + m(x)) + \mathbf{1}_{\{x\in \mathrm{d}y\}}$ on $\mathcal{B}(S)$, where m is a measurable function on S such that $m(x) \geq \overline{q}_x$, then, for each fixed $x \in S$, $\int_S u(y)(q(\mathrm{d}y \mid x, a)/(1 + m(x)) + \mathbf{1}_{\mathrm{d}y}(x))$ is continuous in $a \in A(x)$ (thus, so is $\int_S u(y)q(\mathrm{d}y \mid x, a)$), owing to Lemma 8.3.7 of [19]. This fact is used, for instance, in the proof of Theorem 1 below.

Under Conditions 1, 3, and 4, having in mind the discussion made right above, the fact that a w'-bounded function is also w-bounded by Condition 3(a), and the measurable selection theorem [18, Proposition D.5], we legitimately define the operator T mapping the space of w'-bounded functions on S to itself by

$$T \circ u(x) := \inf_{a\in A(x)}\left\{\frac{c_0(x, a)}{1 + m(x)} + \int_S u(y)\left(\frac{q(\mathrm{d}y \mid x, a)}{1 + m(x)} + \mathbf{1}_{\mathrm{d}y}(x)\right)\right\}, \qquad x \in S, \tag{3}$$

where u is a w'-bounded function on S and m is a measurable function satisfying $m(x) \geq \overline{q}_x$.

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

Theorem 1. *Suppose that Conditions 1, 3, and 4 are satisfied. Then the following assertions hold.*

(a) *The Bellman equation*

$$0 = \inf_{a\in A(x)} \left\{ c_0(x,a) + \int_S v(y) q(\mathrm{d}y \mid x,a) \right\}, \qquad x \in S, \tag{4}$$

admits a w'-bounded solution, say u^*, which is given by the value iteration procedure $u^*(x) := \lim_{n\to\infty} u^{(n)}(x)$, where

$$u^{(0)}(x) := \frac{M'}{\rho'} w'(x),$$
$$u^{(n+1)}(x) := T \circ u^{(n)}(x), \qquad n = 0, 1, \ldots,$$

with the operator T being defined by (3). (We simply observe, which can be shown by induction and the measurable selection theorem [18, Proposition D.5], that, for each $n = 0, 1, \ldots$, $u^{(n)}$ is measurable and w'-bounded on S, so that the proposed value iteration procedure is well defined.)

(b) *The Bellman function u^* defined in part (a) satisfies*

$$u^*(x) = \inf_{\pi\in\Pi_{\mathrm{H}}} V(x,\pi,c_0), \qquad x \in S,$$

and is the unique solution to (4) out of the class of w'-bounded functions on S.

(c) *There is a deterministic stationary optimal policy φ^* for problem (1), which can be taken as any (there exists at least one) measurable mapping $\varphi^*\colon S \to A$ providing the minimizer in the Bellman equation (1), i.e.*

$$\inf_{a\in A(x)} \left\{ c_0(x,a) + \int_S u^*(y) q(\mathrm{d}y \mid x,a) \right\}$$
$$= c_0(x,\varphi^*(x)) + \int_S u^*(y) q(\mathrm{d}y \mid x,\varphi^*(x)), \qquad x \in S,$$

gives a deterministic stationary optimal policy.

(d) *If, in addition, Condition 2(a) is satisfied then the Bellman function u^* coming from part (a) and the deterministic stationary optimal policy φ^* coming from part (c) satisfy $\int_S u^*(x)\gamma(\mathrm{d}x) = \inf_{\pi\in\Pi_{\mathrm{H}}} V(\gamma,\pi,c_0) = V(\gamma,\varphi^*,c_0)$.*

Proof. (a) Firstly we show that the sequence $\{u^{(n)}\}$ is decreasing. Indeed, for each $x \in S$, we have

$$\begin{aligned} u^{(1)}(x) &= \inf_{a\in A(x)} \left\{ \frac{c_0(x,a)}{1+m(x)} + \int_S \frac{M'}{\rho'} w'(y) \left(\frac{q(\mathrm{d}y \mid x,a)}{1+m(x)} + \mathbf{1}_{\mathrm{d}y}(x) \right) \right\} \\ &\le \inf_{a\in A(x)} \left\{ \frac{c_0(x,a)}{1+m(x)} + \frac{M'}{\rho'} \frac{-\rho' w'(x)}{1+m(x)} + w'(x)\frac{M'}{\rho'} \right\} \\ &\le \frac{M'w'(x)}{1+m(x)} - \frac{M'w'(x)}{1+m(x)} + w'(x)\frac{M'}{\rho'} \\ &= w'(x)\frac{M'}{\rho'} \\ &= u^{(0)}(x), \end{aligned}$$

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

where the first and second inequalities are due to Condition 3. This, together with the fact that the operator T is monotonic (increasing), shows that, for each $x \in S$,

$$u^{(n+1)}(x) = T^{n+1} \circ u^{(0)}(x) = T^n \circ T \circ u^{(0)}(x) = T^n \circ u^{(1)}(x) \le T^n \circ u^{(0)}(x) = u^{(n)}(x)$$

holds for each $n = 0, 1, \dots$. Thus, the sequence $\{u^{(n)}\}$ is decreasing. Secondly, we observe that, for each $n = 0, 1, \dots, |u^{(n)}(x)| \le M'w'(x)/\rho'$, $x \in S$. Indeed, this is trivially true for the case $n = 0$. Suppose that the claim holds for $n = m$. Then $u^{(m+1)}(x) \le M'w'(x)/\rho'$ as the sequence $\{u^{(n)}\}$ decreases in n. Moreover, we have

$$\begin{aligned} u^{(m+1)}(x) &= \inf_{a \in A(x)} \left\{ \frac{c_0(x,a)}{1+m(x)} + \int_S u^{(m)}(y) \left(\frac{q(\mathrm{d}y \mid x,a)}{1+m(x)} + \mathbf{1}_{\mathrm{d}y}(x) \right) \right\} \\ &\ge \inf_{a \in A(x)} \left\{ \frac{c_0(x,a)}{1+m(x)} - \int_S \frac{M'}{\rho'} w'(y) \left(\frac{q(\mathrm{d}y \mid x,a)}{1+m(x)} + \mathbf{1}_{\mathrm{d}y}(x) \right) \right\} \\ &\ge \inf_{a \in A(x)} \left\{ \frac{c_0(x,a)}{1+m(x)} + \frac{M'}{\rho'} \frac{\rho' w'(x)}{1+m(x)} - w'(x) \frac{M'}{\rho'} \right\} \\ &\ge -\frac{M'w'(x)}{1+m(x)} + \frac{M'w'(x)}{1+m(x)} - w'(x)\frac{M'}{\rho'} \\ &= -w'(x)\frac{M'}{\rho'}. \end{aligned}$$

Hence, $|u^{(m+1)}(x)| \le w'(x)M'/\rho'$, and the claim follows from the induction. Now it follows from this and the monotone convergence that there exists a w'-bounded measurable function u^* on S such that $\lim_{n\to\infty} u^{(n)}(x) = u^*(x)$ for all $x \in S$. Thirdly, we show that the function u^* solves the Bellman equation (4), which is more conveniently written as $u^*(x) = T \circ u^*(x)$, $x \in S$. Indeed, for any fixed $x \in S$, by its definition, $T \circ u^*(x) \le T \circ u^{(n)}(x) = u^{(n+1)}(x)$, so that upon passing to the limit as $n \to \infty$ we have $T \circ u^*(x) \le u^*(x)$. For the opposite direction, note that, for each $a \in A(x)$ and $n = 0, 1, \dots$,

$$u^{(n+1)}(x) = T \circ u^{(n)}(x) \le \frac{c_0(x,a)}{1+m(x)} + \int_S u^{(n)}(y) \left(\frac{q(\mathrm{d}y \mid x,a)}{1+m(x)} + \mathbf{1}_{\mathrm{d}y}(x) \right),$$

so that, by legally (using Lebesgue's dominated convergence theorem) passing to the limit as $n \to \infty$ and then taking the infimum with respect to $a \in A(x)$, we obtain $u^*(x) \le T \circ u^*(x)$. Hence, $u^*(x) = T \circ u^*(x)$ holds.

(b) Let a policy $\pi \in \Pi_{\mathrm{H}}$ and a w'-bounded measurable function u on S be arbitrarily fixed. Dynkin's formula obtained in Lemma 3 is applicable to the function u. Now, under the conditions of this statement, we can legally add to both sides of Dynkin's formula given in Lemma 3 the term $\mathrm{E}_x^\pi[\int_0^t \int_A c_0(\xi_v, a)\pi(\mathrm{d}a \mid \omega, v)\,\mathrm{d}v]$ (which is w'-bounded for any $t \ge 0$; see the discussion below Condition 3) and passing to the limit as $t \to \infty$ we obtain

$$\begin{aligned} &\mathrm{E}_x^\pi\left[\int_0^\infty \int_A c_0(\xi_t, a)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\right] + \lim_{t\to\infty} \mathrm{E}_x^\pi[u(\xi_t)] - u(x) \\ &\quad = \mathrm{E}_x^\pi\left[\int_0^\infty \int_A \pi(\mathrm{d}a \mid \omega, t) \left\{ \int_S q(\mathrm{d}y \mid \xi_t, a)u(y) + c(\xi_t, a) \right\} \mathrm{d}t\right]. \end{aligned}$$

Since the function u is w'-bounded, we have

$$0 \le \lim_{t\to\infty} \mathrm{E}_x^\pi[|u(\xi_t)|] \le \lim_{t\to\infty} \sup_{x\in X} \frac{|u(x)|}{w'(x)} \mathrm{E}_x^\pi[w'(\xi_t)] \le \lim_{t\to\infty} \sup_{x\in X} \frac{|u(x)|}{w'(x)} L' \,\mathrm{e}^{-\rho t} w(x) = 0,$$

where the last inequality follows by Condition 3 and Lemma 1. Therefore, from the equality derived previously based on Dynkin's formula, we have

$$V(x,\pi,c_0)=u(x)+\mathrm{E}_x^{\pi}\left[\int_0^{\infty}\int_A \pi(\mathrm{d}a\mid\omega,t)\left\{\int_S q(\mathrm{d}y\mid\xi_t,a)u(y)+c(\xi_t,a)\right\}\mathrm{d}t\right]. \quad (5)$$

Now, by replacing u in the above equality with the w'-bounded function u^* from part (a), which satisfies the Bellman equation, we obtain

$$V(x,\pi,c_0)\geq u^*(x) \quad (6)$$

for any $\pi\in\Pi_{\mathrm{H}}$. On the other hand, if we take any measurable selector φ^* such that

$$0=\inf_{a\in A(x)}\left\{c_0(x,a)+\int_S q(\mathrm{d}y\mid x,a)u^*(y)\right\}=c_0(x,\varphi^*(x))+\int_S q(\mathrm{d}y\mid x,\varphi^*(x))u^*(y),$$

whose existence is guaranteed due to the fact that $c_0(x,a)+\int_S q(\mathrm{d}y\mid x,a)u^*(y)$ is lower semicontinuous under the conditions of the statement, Δ is measurable, and the measurable selection theorem given in Proposition D.5 of [18], then by (5), with π replaced by φ^*, we have $V(x,\varphi^*,c_0)=u^*(x)$, $x\in S$. This and (6) lead to $V(x,\varphi^*,c_0)=u^*(x)=\inf_{\pi\in\Pi_{\mathrm{H}}}V(x,\pi,c_0)$. This also proves part (c) of the theorem. As for the uniqueness out of the class of w'-bounded functions on S, one only needs to note that if we replace u in (5) with any other w'-bounded solution, say v^*, to the Bellman equation, then the above reasoning can be applied again to give $u^*(x)=\inf_{\pi\in\Pi_{\mathrm{H}}}V(x,\pi,c_0)=v^*(x)$.

(c) This part has been incidentally proved in the proof of part (b).

(d) The statement follows from the fact that

$$\int_S u^*(x)\gamma(\mathrm{d}x)=\int_S V(x,\varphi^*,c_0)\gamma(\mathrm{d}x)\leq\int_S V(x,\pi,c_0)\gamma(\mathrm{d}x)$$

for each $\pi\in\Pi_{\mathrm{H}}$.

Theorem 1 is about the dynamic programming for the absorbing CTMDP with a total cost criterion, for which the expected long-run average cost is identically equal to zero. In this case, corresponding to the same policy, one can also view the expected total cost as the bias, and the dynamic programming approach for the bias optimality has been considered in, for instance, [40], which, however, imposes restrictive conditions that are generally not satisfied by the absorbing model under consideration; see Assumption C therein.

In the proof of part (b) of Theorem 1, we have indeed incidentally established the following statement, which will be used in Section 5.

Lemma 4. *Suppose that Conditions 1 and 3 are satisfied. Let u be a w'-bounded measurable function on S. Then the following two assertions hold.*

(a) *If $0\geq\int_A\pi(\mathrm{d}a\mid x)\{c_0(x,a)+\int_S q(\mathrm{d}y\mid x,a)u(y)\}$ for all $x\in S$, where π is a stationary policy and u is a w'-bounded measurable function on S, then $V(x,\pi,c_0)\leq u(x)$, $x\in S$.*

(b) *If $0\leq\int_A\pi(\mathrm{d}a\mid x)\{c_0(x,a)+\int_S q(\mathrm{d}y\mid x,a)u(y)\}$ for all $x\in S$, where π is a stationary policy and u is a w'-bounded measurable function on S, then $V(x,\pi,c_0)\geq u(x)$, $x\in S$.*

Proof. The statement follows from (5), whose validity only requires Conditions 1 and 3.

We end this section with an auxiliary statement, which is needed in Section 4, for example.

Lemma 5. *Suppose that Conditions 1 and 2(b) are satisfied. Moreover,* $c_0(x,a) \geq 0$. *Then, for each stationary policy* π, *the function* $V(x,\pi,c_0)$ *solves the equation*

$$0 = \int_A \pi(\mathrm{d}a \mid x)\bigg\{c_0(x,a) + \int_S q(\mathrm{d}y \mid x,a)V(y,\pi,c_0)\bigg\}.$$

Proof. Let a stationary policy π be fixed. Then the statement follows from Theorem 3.1 of [32].

4. Convex analytic approach for constrained CTMDPs

In what follows, we focus on the constrained CTMDP problem (2), for which we assume that the consistency holds, i.e. there exists at least one feasible policy for problem (2).

4.1. Occupation measures and optimality of stationary policies

We start this subsection with the following definition of occupation measures (assuming Conditions 1 and 2(a)).

Definition 2. The occupation measure of a policy $\pi \in \Pi_{\mathrm{H}}$ is a measure η^π on $\mathcal{B}(S \times A)$ concentrated on K, defined by

$$\eta^\pi(\Gamma_S,\Gamma_A) := \mathrm{E}_\gamma^\pi\bigg[\int_0^\infty \mathbf{1}_{\{\Gamma_S \cap X\}}(\xi_t)\pi(\Gamma_A \mid \omega,t)\,\mathrm{d}t\bigg], \qquad \Gamma_S \in \mathcal{B}(S),\ \Gamma_A \in \mathcal{B}(A).$$

We denote by $\mathcal{D} := \{\eta^\pi : \pi \in \Pi_{\mathrm{H}}\}$ the space of all occupation measures. Evidently, it holds that

$$V(\gamma,\pi,u) = \int_K u(x,a)\eta^\pi(\mathrm{d}x,\mathrm{d}a) \quad \text{for all } \pi \in \Pi_{\mathrm{H}}$$

for any w-bounded measurable function u. This fact is used throughout the paper without reference. Under Conditions 1 and 2(a), it follows from Lemma 2 that, for each policy $\pi \in \Pi_{\mathrm{H}}$,

$$\eta^\pi(S,A) = \eta^\pi(X,A) \leq \int_X w(x)\eta^\pi(\mathrm{d}x,A) = \mathrm{E}_\gamma^\pi\bigg[\int_0^\infty w(\xi_t)\,\mathrm{d}t\bigg] \leq \frac{M}{\rho}\int_S w(x)\gamma(\mathrm{d}x) < \infty; \tag{7}$$

in other words, under each policy, the expected absorbing time is finite. This justifies the use of the term ‘absorbing CTMDPs’.

The next statement characterizes the elements of the space $\mathcal{D}$.

Theorem 2. *Suppose that Conditions 1 and 2(a) are satisfied. Then the following assertions hold.*

(a) *The space* $\mathcal{D}$ *is convex, and a measure* η *on* $S \times A$ *concentrated on* K *is in* $\mathcal{D}$ *(i.e.* η *is an occupation measure for some policy) if and only if it satisfies the two relations*

$$0 = \gamma(\Gamma_S \cap X) + \int_K q(\Gamma_S \cap X \mid y,a)\eta(\mathrm{d}y,\mathrm{d}a), \qquad \Gamma_S \in \mathcal{B}(S), \tag{8}$$

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

and

$$\int_S w(y)\eta(\mathrm{d}y, A) \leq \frac{M}{\rho} \int_S w(x)\gamma(\mathrm{d}x) < \infty. \tag{9}$$

(b) *For each policy* $\pi \in \Pi_{\mathrm{H}}$, *there is a stationary policy* π' *such that* $\eta^\pi(\mathrm{d}x, \mathrm{d}a) = \eta^{\pi'}(\mathrm{d}x, \mathrm{d}a)$. *Indeed,* π' *can be taken from the disintegration of* $\eta^\pi(\mathrm{d}x, \mathrm{d}a)$ *with respect to its marginal* $\eta^\pi(\mathrm{d}x, A)$, *i.e.* $\eta^\pi(\mathrm{d}x, \mathrm{d}a) = \pi'(\mathrm{d}a \mid x)\eta^\pi(\mathrm{d}x, A)$.

Proof. (a) The convexity of $\mathcal{D}$ automatically follows from the characterization part, for which we first prove the 'only if' part. We consider an arbitrarily fixed policy $\pi \in \Pi_{\mathrm{H}}$ and its occupation measure $\eta^\pi \in \mathcal{D}$. It follows from (7) that (9) is satisfied by η^π. So we only need verify (8) for η^π. To this end, it is convenient to consider an equivalent definition of the occupation measure via the setwise convergence

$$\eta^\pi(\Gamma_S, \Gamma_A) := \lim_{n\to\infty} \mathrm{E}_\gamma^\pi\bigg[\int_0^\infty \mathrm{e}^{-t/n}\, \mathbf{1}_{\{\Gamma_S \cap X\}}(\xi_t)\pi(\Gamma_A \mid \omega, t)\,\mathrm{d}t\bigg] \tag{10}$$

for each $\Gamma_S \in \mathcal{B}(S)$ and $\Gamma_A \in \mathcal{B}(A)$. Indeed, this definition is legal because of Lévy's monotone convergence theorem and Theorem 4.6.3 of [5]. Note also that the measure defined by

$$\frac{1}{n}\mathrm{E}_\gamma^\pi\bigg[\int_0^\infty \mathrm{e}^{-t/n}\, \mathbf{1}_{\{\Gamma_S \cap X\}}(\xi_t)\pi(\Gamma_A \mid \omega, t)\,\mathrm{d}t\bigg]$$

is in the form of the occupation measure for discounted CTMDPs with the discount factor $1/n$ for each $n = 1, 2, \ldots$, as considered in [31, Definition 3.1] (see also [15, Definition 3.4]). So, by Theorem 3.2 of [31] (see also Theorem 3.5(a) of [15]), we see, for each $\Gamma_S \in \mathcal{B}(S)$,

$$\begin{aligned}
&\frac{1}{n}\mathrm{E}_\gamma^\pi\bigg[\int_0^\infty \mathrm{e}^{-t/n}\, \mathbf{1}_{\{\Gamma_S \cap X\}}(\xi_t)\,\mathrm{d}t\bigg]\\
&\quad = \gamma(\Gamma_S \cap X) + n\int_K q(\Gamma_S \cap X \mid y, a)\frac{1}{n}\mathrm{E}_\gamma^\pi\bigg[\int_0^\infty \mathrm{e}^{-t/n}\, \mathbf{1}_{\{\mathrm{d}y \cap X\}}(\xi_t)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\bigg]\\
&\quad = \gamma(\Gamma_S \cap X) + \int_K q(\Gamma_S \cap X \mid y, a)\,\mathrm{E}_\gamma^\pi\bigg[\int_0^\infty \mathrm{e}^{-t/n}\, \mathbf{1}_{\{\mathrm{d}y \cap X\}}(\xi_t)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\bigg].
\end{aligned}$$

By passing to the limit as $n \to \infty$ on both sides of the above equality, we further obtain

$$\begin{aligned}
0 &= \gamma(\Gamma_S \cap X) + \lim_{n\to\infty}\int_K q(\Gamma_S \cap X \mid y, a)\,\mathrm{E}_\gamma^\pi\bigg[\int_0^\infty \mathrm{e}^{-t/n}\, \mathbf{1}_{\{\mathrm{d}y \cap X\}}(\xi_t)\pi(\mathrm{d}a \mid \omega, t)\,\mathrm{d}t\bigg]\\
&= \gamma(\Gamma_S \cap X) + \int_K q(\Gamma_S \cap X \mid y, a)\eta^\pi(\mathrm{d}y, \mathrm{d}a),
\end{aligned}$$

where the last equality is because of the setwise convergence (10) and Theorem 2.1 of [20]. Thus, (8) is satisfied by η^π, and the 'only if' part is thus proved.

We now prove the 'if' part. Let a measure η on $S \times A$ concentrated on K satisfying (8) and (9) be arbitrarily fixed. By Proposition D.8 of [18] we can take a stationary policy π satisfying $\eta(\mathrm{d}x, \mathrm{d}a) = \pi(\mathrm{d}a \mid x)\eta(\mathrm{d}x, A)$. Now in order to show that $\eta^\pi(\mathrm{d}x, \mathrm{d}a) = \eta(\mathrm{d}x, \mathrm{d}a)$, it suffices to show that $\int_K f(x, a)\eta(\mathrm{d}x, \mathrm{d}a) = \int_K f(x, a)\eta^\pi(\mathrm{d}x, \mathrm{d}a)$ for each nonnegative, bounded, measurable function f on $S \times A$ such that $f(x, a) = 0$ for each $x \in \Delta$ and $a \in A(x)$

as follows. Indeed, we have

$$\begin{aligned}\int_K f(x,a)\eta(\mathrm{d}x,\mathrm{d}a) &= \int_S\int_A f(x,a)\pi(\mathrm{d}a\mid x)\eta(\mathrm{d}x,A)\\ &= \int_S\bigg\{-\int_S\int_A \pi(\mathrm{d}a\mid x)q(\mathrm{d}y\mid x,a)V(y,\pi,f)\bigg\}\eta(\mathrm{d}x,A)\\ &= \int_S V(y,\pi,f)\bigg\{-\int_K q(\mathrm{d}y\mid x,a)\eta(\mathrm{d}x,\mathrm{d}a)\bigg\}\\ &= \int_S V(y,\pi,f)\gamma(\mathrm{d}y),\end{aligned}$$

where the first equality is by the definition of π, the second equality follows from Lemma 5 with f in lieu of c_0 therein, the third equality is because of the Fubini–Tonelli theorem (recalling that $V(y,\pi,f)$ is bounded, having in mind that f is bounded), (7), and (9), and the last equality is due to (8) (recalling that $V(y,\pi,f)=0$ for each $y\in\Delta$ and $a\in A(y)$ because of the definition of the function f). This thus implies that $\int_K f(x,a)\eta(\mathrm{d}x,\mathrm{d}a)=\int_K f(x,a)\eta^{\pi}(\mathrm{d}x,\mathrm{d}a)$, as desired.

(b) This part has been incidentally proved in the proof of the 'if' part of the proof of part (a).

Remark 2. By inspecting the proof of Theorem 2, we see that part (b) still holds if we replace (9) with $\int_S w(y)\eta(\mathrm{d}y,A)<\infty$.

Theorem 2 implies that it suffices to be restricted to the class of stationary policies for the constrained CTMDP problem (2), which can be conveniently rewritten as the following convex program in $\mathcal{D}$:

$$\begin{aligned}&\int_K c_0(x,a)\eta(\mathrm{d}x,\mathrm{d}a)\to\min_{\eta\in\mathcal{D}}\\ \text{s.t.}\quad &\int_K c_j(x,a)\eta(\mathrm{d}x,\mathrm{d}a)\le d_j,\qquad j=1,2,\ldots,N.\end{aligned}\tag{11}$$

This gives rise to the convex analytic approach for the constrained CTMDP problem (2). In what follows, without loss of generality and for simplicity, we directly regard $\eta\in\mathcal{D}$ as measures on $K=\mathbb{K}\setminus(\Delta\times A)$.

In order to obtain the compactness of $\mathcal{D}$ (in an appropriate topology), we need to present further notation and definitions from measure theory [5], [6].

Notation and definitions. Let the Borel spaces K and X be as above, and let a measurable function $f(x)\ge 1$ on X be fixed. We denote by $\mathbb{B}_f(K)$ and $\mathbb{B}_f(X)$ the spaces of measurable functions u on K and X, respectively, with a finite f-norm, i.e. u satisfies

$$\sup_{x\in X}\frac{\sup_{a\in A(x)}|u(x,a)|}{f(x)}<\infty\quad\text{and}\quad\sup_{x\in X}\frac{|u(x)|}{f(x)}<\infty,$$

respectively. Denote by $\mathbb{M}^R(K)$ the space of Radon signed measures on $\mathcal{B}(K)$, i.e. each $\eta\in\mathbb{M}^R(K)$ is a signed measure on $\mathcal{B}(K)$ such that, for every $\Gamma_S\in\mathcal{B}(K)$ and $\varepsilon>0$, there exists a compact set $D_\varepsilon\subseteq\Gamma_S$ that satisfies $|\eta|(\Gamma_S\setminus D_\varepsilon)<\varepsilon$, where here and below $|\eta|$ denotes the total variation of the signed measure η. Nonnegative Radon signed measures are simply called Radon measures, the space of which is denoted by $\mathbb{M}^{R,+}(K)$. We equip $\mathbb{M}^R(K)$ and $\mathbb{M}^{R,+}(K)$ with the usual weak topologies, respectively denoted by $\tau(\mathbb{M}^R(K))$ and $\tau(\mathbb{M}^{R,+}(K))$, which are the weakest topologies such that $\int_K u(x,a)\eta(\mathrm{d}x,\mathrm{d}a)$ is continuous in $\eta \in \mathbb{M}^R(K)$ and $\eta \in \mathbb{M}^{R,+}(K)$, respectively, for each continuous function $u \in \mathbb{B}_1(K)$. A family of signed measures $\tilde{\mathfrak{D}}$ on $\mathcal{B}(K)$ is called uniformly tight if, for every $\varepsilon > 0$, there exists a compact set $D_\varepsilon \subseteq K$ such that $|\tilde{\eta}|(K \setminus D_\varepsilon) < \varepsilon$ for all $\tilde{\eta} \in \tilde{\mathfrak{D}}$. A signed measure $\tilde{\eta}$ on $\mathcal{B}(K)$ is called tight if the singleton $\{\tilde{\eta}\}$ is uniformly tight. It is worthwhile noting that a tight measure on a Borel space endowed with the Borel σ-algebra is Radon by the first paragraph on page 70 and Theorem 7.1.7 of [6].

The next lemma will be used in the proof of Lemma 7 below.

Lemma 6. *Let K be a Borel space endowed with the Borel σ-algebra. The set $\mathbb{M}^{R,+}(K)$ is closed in $(\mathbb{M}^R(K), \tau(\mathbb{M}^R(K)))$.*

Proof. Consider a net $\eta_n \in \mathbb{M}^{R,+}(K)$ such that $\eta_n \to \eta \in \mathbb{M}^R(K)$, where the convergence is in the weak topology introduced above. Now suppose that the statement to be proved is false, i.e. there is some measurable set $\Gamma_1 \subset K$ such that $\eta(\Gamma_1) < 0$. Since η is a Radon signed measure, there is no loss of generality in regarding Γ_1 as a nonempty, compact (and thus closed) set. Then, on the one hand, for every nonnegative, bounded, continuous function f on K, it holds that $0 \leq \lim_{n\to\infty} \int_K f(x,a)\eta_n(\mathrm{d}x, \mathrm{d}a) = \int_K f(x,a)\eta(\mathrm{d}x, \mathrm{d}a)$. On the other hand, we fix $0 < \varepsilon < -\eta(\Gamma_1)$, and again, by the fact that η is Radon, there exists a compact (and thus closed) set $D_\varepsilon \subseteq K \setminus \Gamma_1$, assumed to be nonempty without loss of generality, such that $|\eta|((K \setminus \Gamma_1) \setminus D_\varepsilon) < \varepsilon$. Now, since Γ_1 and D_ε are disjoint closed sets, we refer to Urysohn's lemma stated as Lemma 7.1 in [3] for the existence of a nonnegative, bounded, continuous function f_C on K such that $f_C(x,a) = 1$ for each $(x,a) \in \Gamma_1$, $f(x,a) = 0$ for each $(x,a) \in D_\varepsilon$, and $0 < f_C(x,a) < 1$ for each $(x,a) \in (K \setminus \Gamma_1) \setminus D_\varepsilon$. For this function, we have $\int_K f_C(x,a)\eta(\mathrm{d}x, \mathrm{d}a) = \int_{\Gamma_1} f_C(x,a)\eta(\mathrm{d}x, \mathrm{d}a) + \int_{(K\setminus\Gamma_1)\setminus D_\varepsilon} f_C(x,a)\eta(\mathrm{d}x, \mathrm{d}a) < \eta(\Gamma_1) + \varepsilon < 0$, which is a contradiction. Therefore, $\eta \in \mathbb{M}^{R,+}(K)$, and the statement is proved.

The following version of Prokhorov's theorem is a consequence of Lemma 6 and Theorem 8.6.7 of [6].

Lemma 7. *Let K be a Borel space endowed with the Borel σ-algebra, and let $\tilde{\mathfrak{D}} \subseteq \mathbb{M}^{R,+}(K)$ be uniformly tight and uniformly bounded (i.e. $\sup_{\tilde{\eta}\in\tilde{\mathfrak{D}}} \tilde{\eta}(K) < \infty$). Then $\tilde{\mathfrak{D}}$ is relatively compact (also called precompact) in $(\mathbb{M}^{R,+}(K), \tau(\mathbb{M}^{R,+}(K)))$.*

Proof. By Lemma 7 and the fourth line of page 40 of [1], the closure of $\tilde{D} \subseteq \mathbb{M}^{R,+}(K)$ is compact in $\mathbb{M}^{R,+}(K)$ if and only if it is compact in $\mathbb{M}^R(K)$, which is true by Theorem 8.6.7 of [6].

We also consider the topological spaces introduced in the following definition.

Definition 3. Let a measurable function $f(x) \geq 1$ on $X := S \setminus \Delta$ be fixed.

(a) A Radon signed measure η on K or on X is said to have a finite f-norm if

$$\int_K f(x)|\eta|(\mathrm{d}x, \mathrm{d}a) < \infty \quad \text{or, respectively,} \quad \int_X f(x)|\eta|(\mathrm{d}x) < \infty.$$

The spaces of Radon signed measures on K and on X with finite f-norms are respectively denoted by $\mathbb{M}_f^R(K)$ and $\mathbb{M}_f^R(X)$. The spaces of Radon measures on K and on X with finite f-norms are respectively denoted by $\mathbb{M}_f^{R,+}(K)$ and $\mathbb{M}_f^{R,+}(X)$.

(b) The f-weak topology on $\mathbb{M}_f^R(K)$ is the weakest topology on $\mathbb{M}_f^R(K)$ such that $\int_K u(x,a)\eta(\mathrm{d}x,\mathrm{d}a)$ is continuous in $\eta \in \mathbb{M}_f^R(K)$ for each continuous $u \in \mathbb{B}_f(K)$. This topology is denoted by $\tau(\mathbb{M}_f^R(K))$, and the corresponding convergence is denoted by '$\xrightarrow{f}$.' When $f(x)=1$ for each $x \in X$, we typically omit f from the notation for brevity.

Now let the function $f(x) \geq 1$ on X be further continuous, and consider the topological space $(\mathbb{M}_f^{R,+}(K), \tau(\mathbb{M}_f^{R,+}(K)))$, where $\tau(\mathbb{M}_f^{R,+}(K))$ is the relative topology of $\tau(\mathbb{M}_f^R(K))$ to $\mathbb{M}_f^{R,+}(K)$, and is thus the weakest topology on $\mathbb{M}_f^{R,+}(K)$ such that $\int_K u(x,a)\eta(\mathrm{d}x,\mathrm{d}a)$ is continuous in $\eta \in \mathbb{M}_f^{R,+}(K)$ for each continuous $u \in \mathbb{B}_f(K)$. Then it can be easily shown in Lemma 8 below that $(\mathbb{M}_f^{R,+}(K), \tau(\mathbb{M}_f^{R,+}(K)))$ is homeomorphic with $(\mathbb{M}^{R,+}(K), \tau(\mathbb{M}^{R,+}(K)))$.

Lemma 8. *Let $f(x) \geq 1$ be a fixed, continuous function on X. Then $(\mathbb{M}_f^{R,+}(K), \tau(\mathbb{M}_f^{R,+}(K)))$ is homeomorphic to $(\mathbb{M}^{R,+}(K), \tau(\mathbb{M}^{R,+}(K)))$ with a homeomorphism $Q_f : \mathbb{M}_f^{R,+}(K) \to \mathbb{M}^{R,+}(K)$ defined by (for each $\eta \in \mathbb{M}_f^{R,+}(K)$)*

$$Q_f \circ \eta(\Gamma) := \int_\Gamma f(x)\eta(\mathrm{d}x,\mathrm{d}a), \qquad \Gamma \in \mathcal{B}(K),$$

whose inverse is defined by (for each $\tilde{\eta} \in \mathbb{M}^{R,+}(K)$)

$$Q_f^{-1} \circ \tilde{\eta}(\Gamma) := \int_\Gamma \frac{1}{f(x)}\tilde{\eta}(\mathrm{d}x,\mathrm{d}a), \qquad \Gamma \in \mathcal{B}(K).$$

Proof. We first verify that Q_f is a one-to-one correspondence between $\mathbb{M}_f^{R,+}(K)$ and $\mathbb{M}^{R,+}(K)$. To this end, it suffices to verify that, for arbitrarily fixed $\eta \in \mathbb{M}_f^{R,+}(K)$ and $\tilde{\eta} \in \mathbb{M}^{R,+}(K)$, $f(x)\eta(\mathrm{d}x,\mathrm{d}a)$ and $\tilde{\eta}(\mathrm{d}x,\mathrm{d}a)/f(x)$ are both Radon measures. Since the two measures are both Borel measures, i.e. defined on Borel σ-algebras of Borel spaces, by Theorem 7.1.7 of [6], for them to be Radon measures, it suffices to show that they are tight by the first paragraph on page 70 of [6]; see also the discussion above Lemma 6. Consider now the measure $f(x)\eta(\mathrm{d}x,\mathrm{d}a)$, and let $\varepsilon > 0$ be fixed. Then, by the absolute continuity of integrals, there exists $\delta > 0$ such that, for each measurable set $\Gamma > 0$ with $\eta(\Gamma) < \delta$, it holds that $\int_\Gamma f(x)\eta(\mathrm{d}x,\mathrm{d}a) < \varepsilon$. Since $\eta(\mathrm{d}x,\mathrm{d}a)$ is a Radon measure, there exists a compact set $D \subseteq K$ such that $\eta(K \setminus D) < \delta$, so that $\int_{K\setminus D} f(x)\eta(\mathrm{d}x,\mathrm{d}a) < \varepsilon$. This implies that the measure $f(x)\eta(\mathrm{d}x,\mathrm{d}a)$ is tight, and thus Radon. In exactly the same way, we can show that the measure $\tilde{\eta}(\mathrm{d}x,\mathrm{d}a)/f(x)$ is tight and thus Radon. Therefore, that Q_f is a one-to-one correspondence between $\mathbb{M}_f^{R,+}(K)$ and $\mathbb{M}^{R,+}(K)$ is verified.

We now show that Q_f and Q_f^{-1} are both continuous. Let $\{\eta_n\}$ be a net in $\mathbb{M}_f^{R,+}(K)$ such that $\eta_n \xrightarrow{f} \eta \in \mathbb{M}_f^{R,+}(K)$. Then $Q_f \circ \eta_n =: \tilde{\eta}_n \to \tilde{\eta} =: Q_f \circ \eta$. Indeed, for each continuous function $u \in \mathbb{B}_1(K)$, we see that

$$\begin{aligned}\int_K u(x,a)\tilde{\eta}_n(\mathrm{d}x,\mathrm{d}a) &= \int_K u(x,a)f(x)\eta_n(\mathrm{d}x,\mathrm{d}a)\\ &\to \int_K u(x,a)f(x)\eta(\mathrm{d}x,\mathrm{d}a)\\ &= \int_K u(x,a)\tilde{\eta}(\mathrm{d}x,\mathrm{d}a),\end{aligned}$$

where the convergence follows from the fact that $u(x,a)f(x)$ is a continuous function in $\mathbb{B}_f(K)$.

Similarly, we can show that if $\{\tilde{\eta}_n\}$ is a net in $\mathbb{M}^{R,+}(K)$ such that $\tilde{\eta}_n \to \eta \in \mathbb{M}^{R,+}(K)$, then $Q_f^{-1} \circ \tilde{\eta}_n \xrightarrow{f} Q_f^{-1} \circ \tilde{\eta}$. Thus, the continuity of Q_f and Q_f^{-1} is proved, completing the proof.

We now impose another compactness-continuity condition for the compactness of the space of occupation measures $\mathcal{D}$ in $(\mathbb{M}^R_{w'}(K), \tau(\mathbb{M}^R_{w'}(K)))$ and the existence of a stationary optimal policy for problem (2) as follows.

Condition 5. (a) *The function w from Condition 1 is continuous on X.*

(b) *There exists an increasing sequence of compact sets $K_m \uparrow K$ as $m \uparrow \infty$ such that $\lim_{m\to\infty} \inf_{(x,a)\in K\setminus K_m} w(x)/w'(x) = \infty$, where the function $w'(x) \geq 1$ on X from Condition 3 is assumed to be continuous.*

(c) *For the function w' from part (b),* $\sup_{x\in X} \overline{q}_x / w'(x) < \infty$.

(d) *For each bounded continuous function u on X, $\int_X u(y) q(\mathrm{d}y \mid x, a)$ is continuous in $(x, a) \in K$.*

(e) *For each $i = 0, 1, \ldots, N$, the function $c_i(x, a)$ is lower semicontinuous in $(x, a) \in K$.*

Condition 5(b) implies that $A(x)$ is compact for each $x \in X$ by Lemma 3.10 of [31]. The function w/w' from Condition 5(b) is called a moment by Definition E.7 of [18]. Also, note that, when K is compact, then Condition 5(b) is automatically satisfied because of the convention that the infimum over the empty set is ∞. Finally, Condition 5(b) and (c) imply that $\sup_{x\in X} w'(x)/w(x) < \infty$, a fact that will be used in the proof of Theorem 3(a) below and elsewhere often without special reference.

We can now state the next theorem concerning the solvability of the constrained absorbing CTMDP problem (2).

Theorem 3. *Suppose that Conditions 1, 2(a), and 5(a)–(d) are satisfied. Then the following assertions hold.*

(a) *The space of occupation measures $\mathcal{D}$ is compact in* $(\mathbb{M}^{R,+}_{w'}(K), \tau(\mathbb{M}^{R,+}_{w'}(K)))$.

(b) *If, additionally, Conditions 3(c) and 5(e) are satisfied, then there exists an optimal solution to problem (11), and, thus, there is a (randomized) stationary optimal policy for the constrained absorbing CTMDP problem (2).*

Proof. (a) We first prove that $\mathcal{D}$ is relatively compact in $(\mathbb{M}^{R,+}_{w'}(K), \tau(\mathbb{M}^{R,+}_{w'}(K)))$. Since w' is continuous, as required in Condition 5(b), by Lemma 8, it is equivalent to showing that $\tilde{\mathcal{D}} := \{\tilde{\eta} := Q_{w'} \circ \eta \colon \eta \in \mathcal{D}\}$ is relatively compact in $(\mathbb{M}^{R,+}(K), \tau(\mathbb{M}^{R,+}(K)))$ as follows. By Theorem 2 (see (9)), for the moment function w/w' from Condition 5(b), it holds that $\sup_{\tilde{\eta}\in\tilde{\mathcal{D}}} \int_K w(x)\tilde{\eta}(\mathrm{d}x, \mathrm{d}a)/w'(x) = \sup_{\eta\in\mathcal{D}} \int_K w(x)\eta(\mathrm{d}x, \mathrm{d}a) < \infty$, which, by the generalized version of Proposition E.8 of [18] (from the case of probability measures to that of finite measures), as given in the proof of part (b) of Theorem 3 of [38], implies that $\tilde{\mathcal{D}}$ is uniformly tight. Incidentally, this fact together with the discussion right above Lemma 6 explains why the inequality $\tilde{\mathcal{D}} \subseteq \mathbb{M}^{R,+}(K)$ holds. Again, by Theorem 2 (see (9)) and the discussion following Condition 5, $\sup_{\tilde{\eta}\in\tilde{\mathcal{D}}} \tilde{\eta}(K) < \infty$. Therefore, we refer to Lemma 7 to conclude that $\tilde{\mathcal{D}}$ is relatively compact in $(\mathbb{M}^{R,+}(K), \tau(\mathbb{M}^{R,+}(K)))$.

Next, we show that $\mathcal{D}$ is closed in $(\mathbb{M}^{R,+}_{w'}(K), \tau(\mathbb{M}^{R,+}_{w'}(K)))$. We consider a net $\eta_n \in \mathcal{D}$ such that $\eta_n \xrightarrow{w'} \eta \in \mathbb{M}^{R,+}_{w'}(K)$, where the convergence is in the w'-weak topology. We show

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

that $\eta \in \mathcal{D}$ as follows. Firstly, we verify (9) for η. Indeed, we have

$$\begin{aligned}\int_K w(x)\eta(\mathrm{d}x, \mathrm{d}a) &= \int_K \lim_{m\uparrow\infty} \min\{m, w(x)\}\eta(\mathrm{d}x, \mathrm{d}a)\\ &= \lim_{m\uparrow\infty} \int_K \min\{m, w(x)\}\eta(\mathrm{d}x, \mathrm{d}a)\\ &= \lim_{m\uparrow\infty} \lim_{n\to\infty} \int_K \min\{m, w(x)\}\eta_n(\mathrm{d}x, \mathrm{d}a)\\ &\le \frac{M}{\rho} \int_S w(x)\gamma(\mathrm{d}x),\end{aligned}$$

where in the first equality $\min\{m, w(x)\}$ is a sequence with the index $m = 1, 2, \dots$, the second equality follows by Lévy's monotone convergence theorem, the third equality follows by the continuity of w, as required in Condition 5(a), and the assumption that $\eta_n \overset{w'}{\to} \eta$, with η_n being a net in $\mathcal{D}$, and the last inequality follows by (9) with η_n in lieu of η. Secondly, we verify (8) for η. To this end, let a bounded, continuous, function u on X be arbitrarily fixed. Then we see that

$$\begin{aligned}&\int_X u(x)\left\{\gamma(\mathrm{d}x) + \int_K q(\mathrm{d}x \mid y, a)\eta(\mathrm{d}y, \mathrm{d}a)\right\}\\ &= \int_X u(x)\gamma(\mathrm{d}x) + \int_K \int_X u(x)q(\mathrm{d}x \mid y, a)\eta(\mathrm{d}y, \mathrm{d}a)\\ &= \int_X u(x)\gamma(\mathrm{d}x) + \lim_{n\to\infty} \int_K \int_X u(x)q(\mathrm{d}x \mid y, a)\eta_n(\mathrm{d}y, \mathrm{d}a)\\ &= \lim_{n\to\infty} \int_X u(x)\left\{\gamma(\mathrm{d}x) + \int_K q(\mathrm{d}x \mid y, a)\eta_n(\mathrm{d}y, \mathrm{d}a)\right\}\\ &= 0,\end{aligned}$$

where the second equality follows by Condition 5(c) and (d). This, by the proof of Lemma 6 or, more directly, Lemma 2.3 of [36], implies that $\gamma(\mathrm{d}x) + \int_K q(\mathrm{d}x \mid y, a)\eta(\mathrm{d}y, \mathrm{d}a) = 0$, i.e. (8) is satisfied by η. It remains to refer to Theorem 2 to conclude that $\eta \in \mathcal{D}$, and, thus, $\mathcal{D}$ is closed in $(\mathbb{M}_{w'}^{R,+}(K), \tau(\mathbb{M}_{w'}^{R,+}(K)))$. This, together with the relative compactness of $\mathcal{D}$ as shown at the beginning of this proof, asserts the compactness of $\mathcal{D}$ in $(\mathbb{M}_{w'}^{R,+}(K), \tau(\mathbb{M}_{w'}^{R,+}(K)))$.

(b) Firstly, we note that problem (11) can be written as $\int_K c_0(x, a)\eta(\mathrm{d}x, \mathrm{d}a) \to \min_{\in\mathcal{D}_F}$ with $\mathcal{D}_F := \{\eta^\pi : \pi \in \Pi_\mathrm{F}\} = \{\eta \in \mathcal{D} \colon \int_K c_j(x, a)\eta(\mathrm{d}x, \mathrm{d}a) \le d_j,\ j = 1, 2, \dots, N\}$, where we recall that Π_F is the space of feasible policies for the CTMDP problem (2). Under the additionally imposed Condition 5(e), as in the proof of Theorem 3.11 of [31] based on Lemma A.3 therein, it can be easily proved that, for each $i = 0, 1, \dots, N$, $\int_K c_i(x, a)\eta(\mathrm{d}x, \mathrm{d}a)$ is lower semicontinuous in $\eta \in \mathcal{D}$ equipped with the w'-weak topology. Therefore, the space of feasible occupation measures $\mathcal{D}_F$ is closed in $\mathcal{D}$. This, together with part (a) of this statement, implies that $\mathcal{D}_F \subseteq \mathcal{D}$ is compact, which, again by the lower semicontinuity of $\int_K c_0(x, a)\eta(\mathrm{d}x, \mathrm{d}a)$ in $\eta \in \mathcal{D}_F$, asserts the existence of an optimal solution η^* to problem (11) according to the generalized Weierstrass' theorem stated as Theorem 2.43 of [1]. By Theorem 2, the stationary policy π^* satisfying $\eta^*(\mathrm{d}x, \mathrm{d}a) = \pi^*(\mathrm{d}a \mid x)\eta^*(\mathrm{d}x, A)$ is optimal for the constrained absorbing CTMDP problem (2).

4.2. Optimality of mixed policies

Theorem 1(d) asserts the existence of a deterministic stationary optimal policy for the CTMDP problem (2) with $N = 0$. This result is not covered by Theorem 3(b), which asserts only the existence of (randomized) stationary optimal policies for the CTMDP problem (2) with an arbitrary number, N, of constraints. In this subsection, we give a more detailed characterization of stationary optimal policies for the CTMDP problem (2), which covers Theorem 1(d) as a special case; see the discussion following Definition 4 and Theorem 5 below.

Definition 4. A randomized stationary policy π is said to be $(m+1)$-mixed, where $m = 0, 1, \ldots$, if $\eta^\pi(\mathrm{d}x, \mathrm{d}a) = \sum_{l=0}^{m} b_l \eta^{\varphi_l}(\mathrm{d}x, \mathrm{d}a)$, where the φ_l, $l = 0, 1, \ldots, m$, are deterministic stationary policies, $b_l \geq 0$ for each $l = 0, 1, \ldots, m$, and $\sum_{l=0}^{m} b_l = 1$.

In other words, the occupation measure of an $(m+1)$-mixed (stationary) policy can be expressed as a convex combination of $m+1$ occupation measures generated by deterministic stationary policies. The realization of an $(m+1)$-mixed policy π can be implemented as follows: before the process starts, one selects a deterministic stationary policy φ_l out of the $m+1$ policies with the probability b_l, where the notation is as in Definition 4, and uses it to control the process; see [26, p. 89] for some discussions. Theorem 5 below asserts that, for the absorbing CTMDP problem (2) with N constraints, there exists an $(N+1)$-mixed optimal policy, refining the statement of Theorem 3(b), so that, when $N = 0$, there is a deterministic stationary optimal policy, covering the statement of Theorem 1(d).

We will establish Theorem 5 as a consequence of the following result about the space of performance vectors

$$\mathcal{V} := \Phi(\mathcal{D}) := \left\{ \left(\int_K c_0(x, a) \eta(\mathrm{d}x, \mathrm{d}a), \ldots, \int_K c_N(x, a) \eta(\mathrm{d}x, \mathrm{d}a) \right) : \eta \in \mathcal{D} \right\}$$

of the constrained absorbing CTMDP problem (2). Below, a point $v \in \mathcal{V}$ is said to be generated by a policy π if $v = (V(\gamma, \pi, c_0), \ldots, V(\gamma, \pi, c_N))$.

Theorem 4. *Suppose that Conditions 1, 2(a), 3, 4, and 5(a)–(d) are satisfied. If, additionally, for each $i = 0, 1, \ldots, N$, $c_i(x, a)$ is continuous in $(x, a) \in K$, then the following assertions hold.*

(a) *The space of performance vectors $\mathcal{V}$ is nonempty, convex, and compact in $\mathbb{R}^{N+1}$ equipped with the usual Euclidean topology.*

(b) *Each extreme point of $\mathcal{V}$ is generated by a deterministic stationary policy.*

Proof. (a) Under the conditions of the statement, the mapping Φ defining $\mathcal{V}$ is continuous in η in the w'-weak topology. Since $\mathcal{D}$ is nonempty and convex by Theorem 2, it follows that $\mathcal{V}$ is too, whereas the compactness of $\mathcal{V}$ follows from that of $\mathcal{D}$ (due to Theorem 3), the continuity of Φ, and Theorem 2.34 of [1].

(b) Firstly, we note that, according to part (a) and Corollary 7.66 of [1], there is at least one extreme point of $\mathcal{V}$. So we arbitrarily fix an extreme point v_{ex} of $\mathcal{V}$. We show by induction with respect to the number of constraints N that v_{ex} is generated by a deterministic stationary policy.

Consider $N = 0$. In this case, according to part (a) of this theorem, $\mathcal{V}$ is a closed, bounded interval in $\mathbb{R}$, and there are only two extreme points, which are given by the two endpoints of $\mathcal{V}$. On the other hand, these two endpoints correspond to the optimal values of the problems $V(\gamma, \pi, c_0) \to \min_{\pi \in \Pi_{\mathrm{H}}}$ and $V(\gamma, \pi, c_0) \to \max_{\pi \in \Pi_{\mathrm{H}}}$, respectively, which in turn, by Theorem 1(d) (it can be applied to the maximization problem due to the fact that $c_0(x, a)$ is continuous in $(x, a) \in K$, as required by the conditions of the statement), are generated by deterministic stationary policies. Hence, the fixed extreme point v_{ex} is generated by a deterministic stationary policy.

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

Suppose that the statement holds for the case $N = k-1$, and let us consider $N = k$. It follows that $v_{\rm ex} := (v_{{\rm ex},0}, v_{{\rm ex},1}, \ldots, v_{{\rm ex},k}) \notin \mathcal{V}^\circ$, where $\mathcal{V}^\circ$ stands for the interior of $\mathcal{V}$. So we can refer to the supporting hyperplane theorem, as stated in Proposition 2.4.1 of [4], for the existence of a hyperplane $\mathscr{H} := \{x = (x_0, x_1, \ldots, x_k) \in \mathbb{R}^{k+1}, \sum_{i=0}^k \lambda_i x_i = v^*\}$, where $v^* \in \mathbb{R}$, and at least one $\lambda_i \in \mathbb{R}$, $i = 0, 1, \ldots, N$, is not equal to zero, such that $v_{\rm ex} \in \mathscr{H}$, and $\sum_{i=0}^k \lambda_i v_{{\rm ex},i} \le \sum_{i=0}^k \lambda_i v_i$ holds for each $v \in \mathcal{V}$, i.e. $V(\gamma, \pi^{\rm ex}, \sum_{i=0}^k \lambda_i c_i) = \inf_{\pi \in \Pi_{\rm H}} V(\gamma, \pi, \sum_{i=0}^k \lambda_i c_i)$, where $\pi^{\rm ex}$ is a policy (not necessarily deterministic stationary) that generates $v_{\rm ex}$. Therefore, v^* in the definition of the hyperplane $\mathscr{H}$ is given by

$$v^* := \inf_{\pi \in \Pi_{\rm H}} V\left(\gamma, \pi, \sum_{i=0}^k \lambda_i c_i\right) = V\left(\gamma, \pi^{\rm ex}, \sum_{i=0}^k \lambda_i c_i\right). \tag{12}$$

Below, without loss of generality, we assume that $\lambda_k \neq 0$, for otherwise we may just perform some reordering.

We now define $\mathcal{U} = \mathscr{H} \cap \mathcal{V}$. Since $v^{\rm ex} \in \mathscr{H}$ and $v^{\rm ex} \in \mathcal{V}$, $\mathcal{U} \neq \varnothing$. Since $\mathcal{V}$ is convex and compact, and $\mathscr{H}$ is closed and convex, $\mathcal{U}$ is (nonempty) convex and compact. Thus, $v_{\rm ex}$ is also an extreme point of $\mathcal{U}$ as it is one of $\mathcal{V}$.

As to be shown shortly, it turns out that the space $\mathcal{U}$ coincides with the space of performance vectors $\hat{V}$ of the auxiliary absorbing CTMDP model

$$\{S, A, \hat{A}(x), q({\rm d}y \mid x, a), c_i,\ i = 0, 1, \ldots, k, \gamma, \Delta\},$$

whose validity is yet to be justified below (since we are only concerned with the space of performance vectors, the constraints d_j have been temporarily omitted from consideration and denotation), where $\hat{A}(x)$ is defined by $\hat{A}(x) := A$ for each $x \in \Delta$, and, for each $x \in X$,

$$\begin{aligned}\hat{A}(x) := \bigg\{a^* \in A(x) \colon \sum_{i=0}^k \lambda_i c_i(x, a^*) + \int_S q({\rm d}y \mid x, a^*) u^*(y) \\ = \inf_{a \in A(x)} \bigg\{\sum_{i=0}^k \lambda_i c_i(x, a) + \int_S q({\rm d}y \mid x, a) u^*(y)\bigg\} = 0\bigg\},\end{aligned} \tag{13}$$

with u^* being given by the w'-bounded Bellman function from Theorem 1 with $\sum_{i=0}^k \lambda_i c_i$ in lieu of c_0 therein, and thus satisfying $u^*(x) = \inf_{\pi \in \Pi_{\rm H}} V(x, \pi, \sum_{i=0}^k \lambda_i c_i)$ (here and throughout this proof, in a slight abuse of notation the same symbol u^* is used for the Bellman function even though the cost rate is different from that in Theorem 1; hopefully this does not lead to confusion). Note that, since, under the conditions of the statement, $A(x)$ is compact for each $x \in X$ (see the discussion following Condition 5), the discussion following Condition 4 (noting that $\sup_{x \in X} w(x)/w'(x) < \infty$, as mentioned in the discussion below Condition 5) and the generalized Weierstrass' theorem stated in Theorem 2.43 of [1] imply that $\hat{A}(x)$ is nonempty compact for each $x \in X$, and $\hat{A}(x) \subseteq A$ for each $x \in S$. Therefore, recalling the beginning of Section 2.1, in order for $\{S, A, \hat{A}(x), q({\rm d}y \mid x, a), c_i,\ i = 0, 1, \ldots, k, \gamma, \Delta\}$ to be a valid CTMDP model, we only need further show that $\hat{\mathbb{K}} := \{(x, a) \colon x \in S,\ a \in \hat{A}(x)\}$ is in $\mathscr{B}(S \times A)$ and contains the graph of at least one measurable mapping $\hat{\varphi} \colon S \to A$ such that $\hat{\varphi}(x) \in \hat{A}(x)$ for each $x \in S$.

Indeed, as for the measurability of $\hat{\mathbb{K}}$, we first note that, for each closed set $F \subseteq A$, the set

$$\{x \in X : \hat{A}(x) \cap F \neq \varnothing\} = \left\{x \in X : \inf_{a \in A(x) \cap F} \left\{\sum_{i=0}^{k} \lambda_i c_i(x, a) + \int_S q(\mathrm{d}y \mid x, a) u^*(y)\right\} = 0\right\}$$

is measurable by [21], Theorem 3.1 of [22], and the measurable selection theorem given as Proposition D.5 of [18] validated by the discussions following Conditions 4 and 5. Thus, the multifunction $\hat{A}(x)$ is measurable in the sense of Definition 18.1 of [1]. This in turn validates Theorem 18.6 of [1], which together with the fact that $\hat{A}(x)$ is compact for each $x \in X$ implies that $\hat{K} := \{(x, a) : x \in X, a \in \hat{A}(x)\} \in \mathcal{B}(X \times A)$. Since Δ is measurable in S, it follows that $\hat{\mathbb{K}} = \hat{K} \cup (\Delta \times A) \in \mathcal{B}(S \times A)$. On the other hand, it follows from Theorem 1 (see also its proof) that $\hat{\mathbb{K}}$ contains the graph of at least one measurable mapping $\hat{\varphi} : S \to A$ such that $\hat{\varphi}(x) \in \hat{A}(x)$ for each $x \in S$. Thus, $\{S, A, \hat{A}(x), q(\mathrm{d}y \mid x, a), c_i, i = 0, 1, \ldots, k, \gamma, \Delta\}$ (the functions like c_i will be regarded as their restrictions on $\hat{\mathbb{K}}$) is indeed a valid absorbing CTMDP model, which will be called auxiliary from now on. For this auxiliary model, all the corresponding versions of the conditions of this theorem are satisfied. Indeed, for Condition 5(b), we only need take the compact (in the relative topology) sets $\hat{K}_m := K_m \cap \hat{K}$, whereas the verification of all the other conditions is automatic because the auxiliary model is a submodel of the original model. In particular, the last observation implies that every policy for the auxiliary model is also one for the original model, a fact that is used below without reference.

Now we are ready to show that the space of performance vectors $\hat{\mathcal{V}}$ of the auxiliary model coincides with the space $\mathcal{U} := \mathcal{V} \cap \mathcal{H}$ as defined earlier. Indeed, in one direction, we easily see that $\hat{\mathcal{V}} \subseteq \mathcal{V}$ and $\hat{\mathcal{V}} \subseteq \mathcal{H}$, where the latter inequality follows from the definitions of $\mathcal{H}$ and $\hat{A}(x)$ and Theorems 1 and 2(b). Thus, $\hat{\mathcal{V}} \subseteq \mathcal{U}$. To show the opposite direction of the last inequality, we consider an arbitrarily fixed point $v = (V(\gamma, \pi, c_0), \ldots, V(\gamma, \pi, c_k)) \in \mathcal{U}$, so that π generating v is a policy for the original model, which is assumed to be stationary without loss of generality by Theorem 2, and

$$V\left(\gamma, \pi, \sum_{i=0}^{k} \lambda_i c_i\right) = v^*, \tag{14}$$

where v^* comes from the definition of the hyperplane $\mathcal{H}$ (recalling that v^* satisfies (12)). We now show that v can also be generated by a policy for the auxiliary CTMDP model. Indeed, the measurable set

$$\hat{\Gamma} := \left\{x \in S : \int_A \left(\sum_{i=0}^{k} \lambda_i c_i(x, a) + \int_S q(\mathrm{d}y \mid x, a) u^*(y)\right) \pi(\mathrm{d}a \mid x) > 0\right\},$$

where $u^*(y)$ is the Bellman function as in (13), is null with respect to the measure $\eta^\pi(\mathrm{d}x, A)$, for otherwise we would obtain a contradiction given by

$$0 < \int_X \eta^\pi(\mathrm{d}x, A) \left\{\int_A \pi(\mathrm{d}a \mid x) \sum_{i=0}^{k} \lambda_i c_i(x, a) + \int_X \int_A \pi(\mathrm{d}a \mid x) q(\mathrm{d}y \mid x, a) u^*(y)\right\}$$

$$= V\left(\gamma, \pi, \sum_{i=0}^{k} \lambda_i c_i\right) + \int_X \eta^\pi(\mathrm{d}x, A) \int_S \int_A \pi(\mathrm{d}a \mid x) q(\mathrm{d}y \mid x, a) u^*(y)$$

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

$$
\begin{aligned}
&= V\left(\gamma, \pi, \sum_{i=0}^{k} \lambda_i c_i\right) - \int_X u^*(y)\gamma(\mathrm{d}y) \\
&= v^* - v^* \\
&= 0,
\end{aligned}
$$

where the second equality follows by (8), and the third equality follows by (14), Theorem 1(d) (with $\sum_{i=0}^{k} \lambda_i c_i$ in lieu of c_0), and (13). Therefore, the policy is concentrated on $\hat{A}(x)$ for each $x \in S \setminus \hat{\Gamma}$. Now we define a policy $\hat{\pi}$ for the auxiliary model by $\hat{\pi}(\mathrm{d}a \mid x) := \pi(\mathrm{d}a \mid x)$ for each $x \in S \setminus \hat{\Gamma}$, and $\hat{\pi}(\mathrm{d}a \mid x) := \mathbf{1}_{\{\hat{\varphi}(x)\in \mathrm{d}a\}}$ for each $x \in \hat{\Gamma}$, where $\hat{\varphi}$ is any fixed measurable mapping from S to A such that $\hat{\varphi}(x) \in \hat{A}(x)$ for each $x \in S$, whose existence has been guaranteed earlier in this proof when verifying that $\{S, A, \hat{A}(x), q(\mathrm{d}y \mid x, a), c_i, i = 0, 1, \ldots, k, \gamma, \Delta\}$ is a legal absorbing CTMDP model. The policy $\hat{\pi}$ satisfies $\eta^{\pi}(\mathrm{d}x, A)\hat{\pi}(\mathrm{d}a \mid x) = \eta^{\pi}(\mathrm{d}x, A)\pi(\mathrm{d}a \mid x)$. Now it follows from this and Theorem 2 that $\eta^{\pi}(\mathrm{d}x, \mathrm{d}a) = \eta^{\hat{\pi}}(\mathrm{d}x, \mathrm{d}a)$, so that $v = (V(\gamma, \pi, c_0), \ldots, V(\gamma, \pi, c_k)) = (V(\gamma, \hat{\pi}, c_0), \ldots, V(\gamma, \hat{\pi}, c_k)) \in \hat{\mathcal{V}}$, i.e. $\mathcal{U} \subseteq \hat{\mathcal{V}}$. Thus, $\mathcal{U} = \hat{\mathcal{V}}$ is proved.

Since v_{ex} (the extreme point of $\mathcal{V}$ that was arbitrarily fixed at the beginning of this part of the proof) is an extreme point of $\mathcal{U}$ (explained earlier), so is it one of $\hat{\mathcal{V}}$. On the other hand, for each point $\hat{v} = (\hat{v}_0, \ldots, \hat{v}_k) \in \hat{\mathcal{V}}$, we can always express

$$
\hat{v}_k = \frac{v^* - \sum_{i=0}^{k-1} \lambda_i \hat{v}_i}{\lambda_k} \tag{15}
$$

(v^* and λ_i are from the definition of the hyperplane $\mathcal{H}$ and $\lambda_k \neq 0$ as explained earlier), since $\hat{\mathcal{V}} = \mathcal{U} \subseteq \mathcal{H}$. Therefore, as far as its space of performance vectors is concerned, it suffices to exclude c_k and consider the auxiliary CTMDP model with only k cost rates c_i, $i = 0, 1, \ldots, k-1$, i.e. with only $k-1$ constraints. We denote the space of performance vectors of this auxiliary model with $k-1$ constraints by $\mathcal{V}'$. Then it is easy to see that the point $v'_{\mathrm{ex}} := (v_{\mathrm{ex},0}, \ldots, v_{\mathrm{ex},k-1}) \in \mathcal{V}'$ is an extreme point of $\mathcal{V}'$ (recalling also (15)). Therefore, by the inductive supposition applied to the auxiliary CTMDP model with $k-1$ constraints, the extreme point v'_{ex} is generated by a deterministic stationary policy, which, by (15), also generates v_{ex}. Since v_{ex} is arbitrarily fixed, this completes the induction, completing the proof.

Based on Theorem 4, the Krein–Milman theorem (see Proposition 3.3.1 of [4]), and the Caratheodory theorem [4, pp. 37–38], it is not hard to show the existence of an $(N+2)$-mixed optimal policy for the constrained absorbing CTMDP problem (2), where the number of mixtures $N+2$ comes from the fact that $\mathcal{V} \subseteq \mathbb{R}^{N+1}$ and the Caratheodory theorem. However, this result does not cover Theorem 1(d) for the case of $N = 0$. In order to prove the more refined Theorem 5 below, we need the following lemma.

Lemma 9. *Let f be a concave function on $\mathcal{V}$, where $\mathcal{V}$ is a nonempty convex and compact set in $\mathbb{R}^{N+1}$ with N being a positive integer, and let $\overline{\mathcal{H}}_j := \{v = (v_0, \ldots, v_N) \in \mathbb{R}^{N+1} : v_j \leq d_j\}$, $j = 1, \ldots, N$, where the d_j, $j = 1, 2, \ldots, N$, are constants. Consider the optimization problem (assumed to be consistent)*

$$
\begin{aligned}
& f(v) \to \min_{v \in \mathcal{V}} \\
\text{s.t.} \quad & v \in \overline{\mathcal{H}}_j, \qquad j = 1, 2, \ldots, N.
\end{aligned} \tag{16}
$$

Then there is an optimal solution v^{opt} to problem (16) such that $v^{\mathrm{opt}} = \sum_{l=0}^{N} b_l v^l_{\mathrm{ex}}$, where $\sum_{l=0}^{N} b_l = 1$, $b_l \geq 0$ for each $l = 0, 1, \ldots, N$, and v^l_{ex} (the superscript l does not mean the power) is an extreme point of $\mathcal{V}$.

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

Proof. Since problem (16) is consistent, it follows from the conditions on $\mathcal{V}$ that $\mathcal{V}_F := \mathcal{V} \cap (\bigcap_{j=1}^N \overline{\mathcal{H}}_j)$ is nonempty, convex, and compact. On the other hand, the concave function f is automatically continuous by Proposition 1.4.6 of [4]. Hence, by Bauer's principle, as given by Corollary 7.70 of [1], we see that there exists an extreme point v^{opt} of $\mathcal{V}_F$ solving problem (16). For $v^{\text{opt}} = (v_0^{\text{opt}}, \dots, v_N^{\text{opt}})$, it holds that $v^{\text{opt}} \in \mathcal{V} \cap (\bigcap_{j\in\Lambda} \mathcal{H}_j) \cap (\bigcap_{j\notin\Lambda} \mathcal{H}_j^\circ)$, where, for each $j = 1, 2, \dots, N$, $\mathcal{H}_j^\circ := \{v = (v_0, \dots, v_N) \in \mathbb{R}^{N+1} : v_j < d_j\}$, $\mathcal{H}_j := \overline{\mathcal{H}}_j \setminus \mathcal{H}_j^\circ$, the possibly empty index set Λ is given by $\Lambda := \{j = 1, \dots, N : v_j^{\text{opt}} = d_j\}$, and the intersection over the empty index set is by convention regarded as the universal set. Since v^{opt} is an extreme point of $\mathcal{V}$, so is it one of the convex subset $(\mathcal{V} \cap (\bigcap_{j\in\Lambda} \mathcal{H}_j)) \cap (\bigcap_{j\notin\Lambda} \mathcal{H}_j^\circ) \ni v^{\text{opt}}$ of $\mathcal{V}$. Therefore, by Statement (5.7) of [8], v^{opt} is also an extreme point of $\mathcal{V} \cap (\bigcap_{j\in\Lambda} \mathcal{H}_j)$, which, by [8, Main Theorem], together with the fact that Λ has cardinality not bigger than N, in turn implies that v^{opt} can be expressed as the convex combination of $N+1$ extreme points of $\mathcal{V}$.

Now we are in position to state and prove the existence of an $(N+1)$-mixed optimal policy for problem (2).

Theorem 5. *Suppose that Conditions 1, 2(a), 3, 4, and 5(a)–(d) are satisfied. If, additionally, for each $i = 0, 1, \dots, N$, $c_i(x, a)$ is continuous in $(x, a) \in K$, then there exists an $(N+1)$-mixed optimal policy for the constrained absorbing CTMDP problem (2), where N is the number of constraints.*

Proof. By Theorem 4 and the assumption made at the beginning of Section 4 that (2) is consistent, we see that the space of performance vectors $\mathcal{V}$ satisfies the conditions in Lemma 9. Moreover, the optimal value of problem (2) is given by that of problem (16) with the function f being defined by (for each $v = (v_0, \dots, v_N) \in \mathcal{V}$)

$$f(v) = v_0, \tag{17}$$

which is obviously concave (indeed linear). Therefore, we can refer to Lemma 9 for an optimal solution v^{opt} of problem (16) with f given by (17) such that $v^{\text{opt}} = \sum_{l=0}^N b_l v_{\text{ex}}^l$, where, for each $l = 0, 1, \dots, N$, $b_l \geq 0$ and v_{ex}^l (the superscript l does not mean the power) is an extreme point of $\mathcal{V}$, and $\sum_{l=0}^N b_l = 1$. By Theorem 4 we obtain $N+1$ deterministic stationary policies φ_l, $l = 0, 1, \dots, N$, generating v_{ex}^l, $l = 0, 1, \dots, N$, respectively. Now consider the measure on K defined by $\eta^{\text{opt}}(\mathrm{d}x, \mathrm{d}a) := \sum_{l=0}^N b_l \eta^{\varphi_l}(\mathrm{d}x, \mathrm{d}a)$. By Theorem 2(a), we see that $\eta^{\text{opt}}(\mathrm{d}x, \mathrm{d}a)$ is an occupation measure, and, by Theorem 2(b), there is a stationary policy π^{opt} generating the measure $\eta^{\text{opt}}(\mathrm{d}x, \mathrm{d}a)$. It is evident that the policy π^{opt} is the required $(N+1)$-mixed optimal policy for problem (2).

5. Linear programming formulation and strong duality

In this section we will view problem (11) as a (primal) linear program, for which some duality results will be derived. The motivation behind this comes from the fact that, compared to the primal program, the dual program can sometimes be easier to solve. However, in the infinite-dimensional case such as we are concerned with, generally speaking, the values of the primal and dual programs might be different. In what follows we provide conditions for the absence of such duality gaps. To this end, we need to select appropriate linear spaces first. So we introduce the following additional notation.

Notation. Let $w(x) \geq 1$ and $w'(x) \geq 1$ be measurable (later on they are further assumed to be continuous) functions on X. We denote by $\mathbb{M}_{w'}(X)$ the space of signed (not necessarily

Radon) measures on X with a finite w'-norm, i.e. $\nu \in \mathbb{M}_{w'}(X)$ satisfies $\int_X w'(x)|\nu|(\mathrm{d}x) < \infty$. The w'-weak topology on this space is denoted by $\tau(\mathbb{M}_{w'}(X))$, which is generated by the class of continuous functions $u \in \mathbb{B}_{w'}(X)$. Furthermore, we introduce

$$\mathcal{X} := \mathbb{M}_w^R(K) \times \mathbb{R}^N, \qquad \mathcal{Y} := \mathbb{B}_w(K) \times \mathbb{R}^N,$$
$$\mathcal{Z} := \mathbb{M}_{w'}(X) \times \mathbb{R}^N, \qquad \mathcal{W} := \mathbb{B}_{w'}(X) \times \mathbb{R}^N.$$

In what follows, it is standard practice and thus will not be repeated to define elements of the above spaces by

$$\underline{x} := (\eta(\mathrm{d}x, \mathrm{d}a),\ x_1, \dots, x_N) \in \mathcal{X}, \qquad \underline{y} := (g(x, a),\ y_1, \dots, y_N) \in \mathcal{Y},$$
$$\underline{z} := (\nu(\mathrm{d}x),\ z_1, \dots, z_N) \in \mathcal{Z}, \qquad \underline{w} := (h(x),\ w_1, \dots, w_N) \in \mathcal{W},$$

respectively. We introduce, for each $\underline{x}$, $\underline{y}$, $\underline{z}$, $\underline{w}$,

$$\langle \underline{x}, \underline{y} \rangle_1 := \int_K g(x, a)\eta(\mathrm{d}x, \mathrm{d}a) + \sum_{j=1}^N x_j y_j, \qquad \langle \underline{z}, \underline{w} \rangle_2 := \int_X h(x)\nu(\mathrm{d}x) + \sum_{j=1}^N z_j w_j. \quad (18)$$

Below we simply use $\langle \cdot, \cdot \rangle$ for both $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, as the context will always make it clear which one we mean.

The next lemma collects some observations about the above four spaces, which are needed for the linear program formulations of problem (11). In its statement, we freely use (i.e. without explicit references) the terminologies introduced in Chapter 6 of [18] and Chapter 12 of [19].

Lemma 10. *Suppose that Conditions 2(a), 3, and 5(b) (where the continuity of w' is not needed for the moment) are satisfied. Then the following assertions hold.*

(a) *The spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ are all linear spaces.*

(b) *$(\mathcal{X}, \mathcal{Y})$ and $(\mathcal{Z}, \mathcal{W})$ with the bilinear forms defined by (18) are dual pairs.*

(c) *The set $\mathcal{K} := \{\underline{x} \in \mathcal{X} \colon \eta \in \mathbb{M}_w^{R,+}(K),\ x_j \geq 0,\ j = 1, 2, \dots, N\}$ is a positive cone in $\mathcal{X}$ with the dual cone $\mathcal{K}^* := \{\underline{y} \in \mathcal{Y} \colon \langle \underline{x}, \underline{y} \rangle \geq 0$ for all $\underline{x} \in \mathcal{K}\} = \{\underline{y} \in \mathcal{Y} \colon g(x, a) \geq 0$ for all $(x, a) \in K,\ y_j \geq 0,\ j = 1, 2, \dots, N\}$.*

(d) *The mapping G from $\mathcal{X}$ to $\mathcal{Z}$ defined by*

$$G(\underline{x}) := \Big(-\int_K q(\mathrm{d}x \mid y, a)\eta(\mathrm{d}y, \mathrm{d}a), \int_K c_1(x, a)\eta(\mathrm{d}x, \mathrm{d}a) + x_1, \dots, \int_K c_N(x, a)\eta(\mathrm{d}x, \mathrm{d}a) + x_N\Big)$$

has its adjoint G^* given by

$$G^*(\underline{w}) := \Big(-\int_X h(y)q(\mathrm{d}y \mid x, a) + \sum_{j=1}^N w_j c_j(x, a), w_1, \dots, w_N\Big).$$

(e) *The mapping G defined in part (d) is $\tau(\mathcal{X}, \mathcal{Y}) - \tau(\mathcal{Z}, \mathcal{W})$ continuous (also called weakly continuous), where $\tau(\mathcal{X}, \mathcal{Y})$ denotes the weakest topology on $\mathcal{X}$ such that $\langle \cdot, \underline{y} \rangle$ is continuous on $\mathcal{X}$ for each fixed $\underline{y} \in \mathcal{Y}$, and the topology $\tau(\mathcal{Z}, \mathcal{W})$ on $\mathcal{Z}$ is defined similarly.*

Proof. For part (a), the only thing to be verified is the linearity of $\mathcal{X}$ as that of the other three spaces is evident. It follows from Propositions 19.20, 19.43 and Definition 19.41 of [37] and Theorem 12.4 of [1] that our definition of a (finite) Radon signed measure (from [6]) is equivalent to that given in Definition 19.19 of [37]. Now the linearity of the space $\mathbb{M}_w^R(K)$ follows from Propositions 19.39 and 19.44 of [37], which assert that the space of Radon signed measures is linear, and the obvious fact that the linear combination of Radon signed measures with a finite w-norm is again one with a finite w-norm. The linearity of $\mathcal{X}$ now follows.

Parts (b) and (c) of the statement are obvious.

For part (d), we can directly verify that Equation (6.2.2) of [18], which is the same as Equation (12.2.15) of [19], is satisfied. The involved interchanges of the order of integrations are legal under the conditions of the statement, and especially by the fact that the function w' has a finite w-norm, i.e. $\sup_{x\in X} w'(x)/w(x) < \infty$, according to the discussion immediately following Condition 5(b).

Part (e) now follows from the fact that $G^*(\mathcal{W}) \subseteq \mathcal{Y}$ (as can be easily seen) and Proposition 12.2.5 of [19].

Under Conditions 1, 2(a), 3, and 5(b) (where the continuity of w' is not needed for the moment), problem (11) is equivalent (by Theorem 2, the discussion after its proof, and Remark 2) to the following primal linear program, which is well defined due to Lemma 10 and Chapter 6 of [18]:

$$\langle \underline{x}, \hat{y} \rangle \to \min_{\underline{x}\in\mathcal{X}} \quad \text{s.t.} \quad G(\underline{x}) = \hat{z}; \qquad \underline{x} \in \mathcal{K}, \tag{19}$$

where $\hat{y} := (c_0, 0, \dots, 0) \in \mathcal{Y}$ and $\hat{z} := (\gamma, d_1, \dots, d_N) \in \mathcal{Z}$ are fixed points. Indeed, under Condition 5(b), the function w is a moment (see Definition E.7 of [18]), since the function w/w' is, where $w'(x) \geq 1$ on $x \in X$. This, together with Theorem 2 (see (9)), implies that $\mathcal{D}$ is uniformly tight according to the proof of Theorem 3(b) of [38]. As each $\eta \in \mathcal{D}$ is a measure on $\mathcal{B}(K)$ and satisfies (9), we see that $\mathcal{D} \subseteq \mathbb{M}_w^R(K)$ by the discussion above Lemma 6. Now, again by Chapter 6 of [18], the dual linear program of (19) is given by

$$\langle \hat{z}, \underline{w} \rangle \to \max_{\underline{w}\in\mathcal{W}} \quad \text{s.t.} \quad \hat{y} - G^*(\underline{w}) \in \mathcal{K}^*; \qquad \underline{w} \in \mathcal{W},$$

or, more explicitly and after some rearrangements,

$$\int_X h(x)\gamma(\mathrm{d}x) - \sum_{j=1}^N d_j w_j \to \max_{(h(x), w_1, \dots, w_N)\in\mathcal{W}} \tag{20}$$

$$\text{s.t.} \quad c_0(x, a) + \sum_{j=1}^N w_j c_j(x, a) + \int_X h(y) q(\mathrm{d}y \mid x, a) \geq 0 \text{ for all } (x, a) \in K; \quad w_j \geq 0.$$

Below we denote the values of problems (19) and (20) by inf (PLP(19)) and sup(DLP(20)). We collect some observations in the following remark, where the function w could be understood as a fixed measurable function on X such that $w(x) \geq 1$ for each $x \in X$.

Remark 3. Consider the dual pair of the linear spaces $\mathbb{M}_w^R(K)$ and $\mathbb{B}_w(K)$ with the bilinear form $\langle \eta, g \rangle := \int_K g(x, a)\eta(\mathrm{d}x, \mathrm{d}a)$, where $g \in \mathbb{B}_w(K)$ and $\eta \in \mathbb{M}_w^R(K)$. Denote by $\tau(\mathbb{M}_w^R(K), \mathbb{B}_w(K))$ the weakest topology on $\mathbb{M}_w^R(K)$ such that $\langle \cdot, g \rangle$ is continuous in $\eta \in \mathbb{M}_w^R(K)$ for each fixed $g \in \mathbb{B}_w(K)$. Then, by the discussion on page 211 and Theorem 5.93 of [1], the topology $\tau(\mathbb{M}_w^R(K), \mathbb{B}_w(K))$ is compatible with the bilinear form $\langle \cdot, \cdot \rangle$ (defined earlier in this remark) in the sense of [35, p. 13]. It follows from its definition that this topology is stronger than the w-weak topology $\tau(\mathbb{M}_w^R(K))$. In particular, lower semicontinuous functions and closed sets in the w-weak topology $\tau(\mathbb{M}_w^R(K))$ are automatically closed in the topology $\tau(\mathbb{M}_w^R(K), \mathbb{B}_w(K))$.

Remark 3 and the standard Slater condition (see Condition 6) imposed below will be used to validate the statement of Theorem 17(a) of [35], which plays an essential role in the proof of Theorem 6 below about the strong duality between the primal linear program (19) and its dual program (20).

Condition 6. *There exists a policy* $\pi \in \Pi_{\mathrm{H}}$ *such that* $V(\gamma, \pi, c_j) < d_j,\ j = 1, 2, \ldots, N$.

We are now ready to state the following strong duality theorem.

Theorem 6. *Suppose that Conditions 1, 2(a), 3, 4, 5, and 6 are satisfied. Then the strong duality between the primal linear program (19) and its dual program (20) holds, i.e. both problems (19) and (20) admit optimal solutions, and* $\inf(\mathrm{PLP}(19)) = \sup(\mathrm{DLP}(20))$.

Proof. Similarly to the proof of Theorem 3(a), we can show that $\mathcal{D}$ is w-weakly closed in $\mathbb{M}_w^{R,+}(K)$, which is w-weakly closed in $\mathbb{M}_w^R(K)$, as can be verified in a similar way to the proof of Lemma 6. It is also an easy exercise to show that, for each $i = 0, 1, \ldots, N$, $\int_K c_i(x, a)\eta(\mathrm{d}x, \mathrm{d}a)$ is w-weakly lower semicontinuous in $\eta \in \mathcal{D}$. By Remark 3, these two observations lead to the fact that, for each $i = 0, 1, \ldots, N$, $\int_K c_i(x, a)\eta(\mathrm{d}x, \mathrm{d}a)$ is lower semicontinuous on $\mathcal{D} \subseteq (\mathbb{M}_w^R(K), \tau(\mathbb{M}_w^R(K), \mathbb{B}_w(K)))$, and $\mathcal{D}$ is nonempty convex and closed in $(\mathbb{M}_w^R(K), \tau(\mathbb{M}_w^R(K), \mathbb{B}_w(K)))$, where the nonemptiness is obvious, and the convexity follows from Theorem 2(a). These facts, Condition 6, and Remark 3 allow us to refer to Example 1” and Theorem 17 of [35] for

$$\inf(\mathrm{PLP}(19)) = \sup_{w_j \geq 0,\ j=1,2,\ldots,N} \inf_{\eta \in \mathcal{D}} \left\{ \int_K \eta(\mathrm{d}x, \mathrm{d}a)\left(c_0(x, a) + \sum_{j=1}^N w_j c_j(x, a)\right) - \sum_{j=1}^N w_j d_j \right\}. \tag{21}$$

For arbitrarily fixed $w_1 \geq 0, \ldots, w_N \geq 0$, we claim that

$$\begin{aligned} &\inf_{\eta \in \mathcal{D}} \left\{ \int_K \eta(\mathrm{d}x, \mathrm{d}a)\left(c_0(x, a) + \sum_{j=1}^N w_j c_j(x, a)\right) - \sum_{j=1}^N w_j d_j \right\} \\ &\quad = \sup_{h \in \mathbb{B}_{w'}(X)} \left\{ \int_X \gamma(\mathrm{d}x) h(x) \colon c_0(x, a) + \sum_{j=1}^N w_j c_j(x, a) \right. \\ &\qquad\qquad \left. + \int_X h(y) q(\mathrm{d}y \mid x, a) \geq 0 \text{ for all } (x, a) \in K \right\} - \sum_{j=1}^N w_j d_j. \end{aligned} \tag{22}$$

Indeed, by Theorem 1(d) and the discussion following Theorem 2, we see that

$$\begin{aligned} \inf_{\eta \in \mathcal{D}} \left\{ \int_K \eta(\mathrm{d}x, \mathrm{d}a)\left(c_0(x, a) + \sum_{j=1}^N w_j c_j(x, a)\right) \right\} &= \inf_{\pi \in \Pi_{\mathrm{H}}} V\left(\gamma, \pi, c_0 + \sum_{j=1}^N w_j c_j\right) \\ &= \int_X \gamma(\mathrm{d}x) u^*(x), \end{aligned} \tag{23}$$

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

where $u^* \in \mathbb{B}_{w'}(X)$ is the solution to the Bellman equation (4) with $c_0 + \sum_{j=1}^N w_j c_j$ in lieu of c_0 therein. Therefore, u^* is feasible for the maximization problem given on the right-hand side of (22), for which, we suppose that there is some other feasible $\overline{h}$ satisfying $\int_X \overline{h}(x)\gamma(\mathrm{d}x) > \int_X u^*(x)\gamma(\mathrm{d}x)$. Then there exist some $\overline{x} \in X$ and $\varepsilon > 0$ such that $\overline{h}(\overline{x}) > u^*(\overline{x}) + \varepsilon$. Since $\overline{h}$ is feasible, by Lemma 4, $V(\overline{x}, \pi, c_0 + \sum_{j=1}^N w_j c_j) \geq \overline{h}(\overline{x})$ holds for each stationary policy π. Therefore, $\inf_\pi V(\overline{x}, \pi, c_0 + \sum_{j=1}^N w_j c_j) - \varepsilon > u^*(\overline{x})$, where the infimum is taken over all stationary policies π. However, this contradicts the fact that $u^*(x) = \inf_{\pi \in \Pi_{\mathrm{H}}} V(x, \pi, c_0 + \sum_{j=1}^N w_j c_j)$ for each $x \in X$ given by Theorem 1. Hence, (22) follows. By (22), we see that sup(DLP(20)) coincides with the term on the right-hand side of (21). Thus, it follows from (21) that inf(PLP(19)) = sup(DLP(20)).

The solvability of problem PLP(19) is guaranteed by Theorem 3(b). As for DLP(20), we note that an optimal solution is given by $(u^*, w_1^*, \ldots, w_N^*)$, where $u^* \in \mathbb{B}_{w'}(X)$ is the Bellman function from (23), and $w_1^* \geq 0, \ldots, w_N^* \geq 0$ solves the maximization problem on the right-hand side of (21), whose existence is guaranteed by Theorem 17 of [35]. Thus, the proof is completed.

We finally remark that the constants $w_1, \ldots, w_N$ in the dual linear problem (20) are just the Lagrange multipliers with the Lagrangian being $\int_K \eta(\mathrm{d}x, \mathrm{d}a)(c_0(x, a) + \sum_{j=1}^N w_j c_j(x, a)) - \sum_{j=1}^N w_j d_j$; see (21).

6. Examples

In this section we present two examples to illustrate the verifications of the imposed conditions in this paper. We will not consider the standard Slater condition (Condition 6) since it is not imposed on the primitives; anyway, in practice its verification is typically straightforward.

Example 1. Consider a controlled birth-and-death process with the state space $S = \{0, 1, \ldots, \}$ and absorbing set $\Delta := \{0\}$ (and thus $X := \{1, 2, \ldots\}$). Let A be any arbitrarily fixed compact Borel space, and let $A(x) \equiv A$ for each $x \in S$. So $\mathbb{K} = S \times A$ and $K = X \times A$. The transition rate $q(\mathrm{d}y \mid x, a)$ and the cost rates $c_i(x, a)$, $i = 0, 1, \ldots, N$, satisfy Assumption 1 below.

Assumption 1. (a) $q_x(a) = q(\{x + 1\} \mid x, a) + q(\{x - 1\} \mid x, a)$, $q(\{x + 1\} \mid x, a) \geq 0$, $q(\{x - 1\} \mid x, a) \geq \underline{q} > 0$ *for each* $x > 0$ *and* $a \in A$, *where* $\underline{q}$ *is a constant; and* $\inf_{a \in A, x>0} q(\{x - 1\} \mid x, a)/q(\{x + 1\} \mid x, a) \geq \zeta$, *where* $\zeta > 1$ *is a constant.*

(b) *There exist constants* $1 < \zeta_1 < \zeta$, $1 < \zeta_2 < \zeta/\zeta_1$, $\zeta_1 > \zeta_2$, $C_1 > 0$, *and* $C_2 > 0$ *such that*

$$\sup_{a \in A, x>0} \frac{|c_i(x, a)|}{\zeta_1^x} \leq C_1$$

for each $i = 0, 1, \ldots, N$, *and*

$$\sup_{a \in A, x>0} \frac{q(\{x + 1\} \mid x, a) + q(\{x - 1\} \mid x, a)}{\zeta_2^x} \leq C_2.$$

(c) $q(\{x + 1\} \mid x, a)$, $q(\{x - 1\} \mid x, a)$, *and* $c_i(x, a)$, $i = 0, 1, 2, \ldots, N$, *are all continuous in* $a \in A$ *for each* $x \in X$.

The initial distribution γ satisfies $\sum_{x \in S}(\zeta_1\zeta_2)^x \gamma(x) < \infty$.

Proposition 1. *For Example 1, Conditions 1, 2, 3, 4, and 5 are all satisfied.*

https://doi.org/10.1239/aap/1370870127 Published online by Cambridge University Press

Proof. In this proof, for brevity, we define the birth rate $\lambda(x,a) := q(\{x+1\} \mid x,a)$ and the death rate $\mu(x,a) := q(\{x-1\} \mid x,a)$ $(x > 0)$. We also put $w(x) = (\zeta_1\zeta_2)^x\, \mathbf{1}_X(x)$ and $w'(x) = \zeta_1^x\, \mathbf{1}_X(x)$.

Part (b) of Condition 1 can be directly verified by Assumption 1. For part (a) of Condition 1, we note that, for each $x > 0$,

$$\begin{aligned}
&\lambda(x,a)w(x+1) + \mu(x,a)w(x-1) - (\lambda(x,a)+\mu(x,a))w(x)\\
&\quad = \mu(x,a)(\zeta_1\zeta_2)^x\left\{\frac{\lambda(x,a)}{\mu(x,a)}(\zeta_1\zeta_2) + \frac{1}{\zeta_1\zeta_2} - \frac{\lambda(x,a)}{\mu(x,a)} - 1\right\}\\
&\quad \le \mu(x,a)(\zeta_1\zeta_2)^x\left\{\frac{1}{\zeta}(\zeta_1\zeta_2 - 1) - \left(1 - \frac{1}{\zeta_1\zeta_2}\right)\right\}\\
&\quad = \mu(x,a)(\zeta_1\zeta_2)^x\left\{\frac{\zeta_1\zeta_2}{\zeta}\left(1 - \frac{1}{\zeta_1\zeta_2}\right) - \left(1 - \frac{1}{\zeta_1\zeta_2}\right)\right\}\\
&\quad \le -\underline{q}\left(1 - \frac{1}{\zeta_1\zeta_2}\right)\left(1 - \frac{\zeta_1\zeta_2}{\zeta}\right)w(x).
\end{aligned}$$

Thus, Condition 1(a) is verified since $\underline{q}(1 - 1/\zeta_1\zeta_2)(1 - \zeta_1\zeta_2/\zeta) > 0$.

The verification of Condition 2 is trivial, and so is that of Condition 3(a) and (c). For Condition 3(b), we see that, for each $x > 0$,

$$\begin{aligned}
&\lambda(x,a)w'(x+1) + \mu(x,a)w'(x-1) - (\lambda(x,a)+\mu(x,a))w'(x)\\
&\quad = \mu(x,a)\zeta_1^x\left\{\frac{\lambda(x,a)}{\mu(x,a)}\zeta_1 + \frac{1}{\zeta_1} - \frac{\lambda(x,a)}{\mu(x,a)} - 1\right\}\\
&\quad \le \mu(x,a)\zeta_1^x\left\{\frac{1}{\zeta}(\zeta_1 - 1) - \left(1 - \frac{1}{\zeta_1}\right)\right\}\\
&\quad = \mu(x,a)\zeta_1^x\left\{\frac{\zeta_1}{\zeta}\left(1 - \frac{1}{\zeta_1}\right) - \left(1 - \frac{1}{\zeta_1}\right)\right\}\\
&\quad \le -\underline{q}\left(1 - \frac{1}{\zeta_1}\right)\left(1 - \frac{\zeta_1}{\zeta}\right)w'(x).
\end{aligned}$$

Thus, Condition 3(b) is verified since $\underline{q}(1 - 1/\zeta_1)(1 - \zeta_1/\zeta) > 0$.

We verify Condition 4(a), (b), and (c) using Assumption 1(c), and verification of Condition 4(d) is trivial. We verify Condition 5(a), (c), (d), and (e) trivially, since the state space S is countable. Regarding Condition 5(b), we can take $K_m = \{1, \dots, m\} \times A$.

Remark. The unconstrained version of the model described in Example 1 is studied more carefully in [29], where the authors are restricted to the class of deterministic stationary policies as an initial assumption. Thus, Proposition 1 and the optimality results obtained in the present paper justify their assumption. By the way, Assumption 1 is stronger than the conditions in [29] because there the authors only considered the value function of the underlying absorbing CTMDP, and the existence of an optimal policy was not needed for their studies. Note also that, since we allow N, the number of constraints, to be arbitrarily fixed, the result in [16] is not applicable.

Example 2. Consider an economic entity (a company for instance). The wealth is denoted by $x \in S = (-\infty, \infty)$. When the wealth is negative, the economic entity goes bankrupt. Since we are only interested in the period before bankruptcy, $\Delta = (-\infty, 0)$ is taken as the absorbing set, and, thus, $X = [0, \infty)$. The decision maker decides the amount of wealth to be invested given the current wealth $x \in X$, denoted by $a \in A(x) = [0, x] \subset A = [0, \infty)$. If a is invested given the current wealth x, after an exponentially distributed random time with rate $\lambda > 0$, the wealth is changed to a new state following the uniform distribution $U[x - a, x + a + 1]$ with the density $\mathbf{1}_{[x-a,x+a+1]}(y)/(2a + 1)$. Note that the uniform distribution can be understood as a noninformative prior. Moreover, lethal disasters (financial crisis) reducing the wealth directly to Δ occur with a rate given by a continuous function $\beta(x)$, which is measured by a probability $\mu(\cdot)$ on Δ, where we suppose that $\sup_{x\in X} \beta(x)/(x + 1) < \infty$ and $\beta(x) > 4\lambda$ for all $x \in X$. Therefore, the transition rates are taken to be

$$q(\Gamma_S \mid x, a) = \beta(x)\mu(\Gamma_S \cap \Delta) + \lambda \int_{\Gamma_S\cap[x-a,x+a+1]} \frac{1}{2a+1}\,\mathrm{d}y - (\lambda + \beta(x))\,\mathbf{1}_{\Gamma_S}(x)$$

for $x \in X, a \in A(x)$, and $\Gamma_S \in \mathcal{B}(S)$. Furthermore, the initial distribution is such that $\int_X x^2\gamma(\mathrm{d}x) < \infty$, and the cost rates $c_i(x, a)$, $i = 0, 1, \ldots, N$, are continuous in $(x, a) \in K$ and satisfy $\sup_{x\in X}\{\sup_{a\in A(x)} |c_i(x, a)|/(1 + x)\} < \infty$ for each $i = 0, \ldots, N$, where N is the number of constraints.

Proposition 2. *For Example 2, Conditions 1, 2, 3, 4, and 5 are all satisfied.*

Proof. We put $w(x) = \mathbf{1}_X(x)(1 + x^2)$ and $w'(x) = \mathbf{1}_X(x)(1 + x)$, $x \in S$.

Part (b) of Condition 1 can be easily verified. For Condition 1(a), we see that, for $x \in X$ and $a \in A(x)$,

$$\begin{aligned}\int_S q(\mathrm{d}y \mid x, a)w(y) &= \lambda \int_{x-a}^{x+a+1} \frac{1}{2a+1}(1 + y^2)\,\mathrm{d}y - (\lambda + \beta(x))(1 + x^2)\\ &\le \lambda(1 + 3x^2 + a^2) - (\lambda + \beta(x))(1 + x^2)\\ &\le 4\lambda(1 + x^2) - (\lambda + \beta(x))(1 + x^2)\\ &\le (3\lambda - \beta(x))(1 + x^2)\\ &\le -\lambda(1 + x^2).\end{aligned}$$

Thus, Condition 1(a) is satisfied since $\lambda > 0$.

It is trivial to verify Conditions 2, 3(a), and 3(c). For Condition 3(b), we have, for each $x \in X$ and $a \in A(x)$,

$$\begin{aligned}\int_S q(\mathrm{d}y \mid x, a)w'(y) &= \lambda \int_{x-a}^{x+a+1} \frac{1}{2a+1}(1 + y)\,\mathrm{d}y - (\lambda + \beta(x))(1 + x)\\ &\le \lambda(1 + x) - (\lambda + \beta(x))(1 + x)\\ &\le -\beta(x)(1 + x)\\ &\le -4\lambda(1 + x).\end{aligned}$$

Thus, Condition 3(b) is verified.

We verify Conditions 4 and 5(a), (c), (d), and (e) straightforwardly from the definition of the transition rates, and the fact that $\beta(x)$ and $c_i(x, a)$, $i = 0, 1, \ldots, N$, are continuous on K. For Condition 5(b), we can take $K_m = \{(x, a) : x \in [0, m],\ a \in A(x)\}$.

Remark. Since the state space is uncountable, and the transition rate is unbounded, the previous works on this topic (see [16], [28], [33], and the references therein) seem not to cover Example 2.

7. Conclusion

In this paper we developed both the dynamic programming approach and the convex analytic approach for unconstrained and constrained absorbing CTMDPs with total undiscounted cost criteria. Specifically, we obtained the Bellman equation for the unconstrained CTMDPs and proved the existence of a deterministic stationary optimal policy; for the constrained CTMDPs, we defined the space of occupation measures and showed it to be convex and compact, and characterized the extreme points of the convex space of performance vectors and showed them to be generated by deterministic stationary policies, leading to the existence of an $(N+1)$-mixed optimal policy, with N being the number of constraints. Finally, we introduced appropriate dual pairs to formulate the CTMDPs as linear programs, and showed the strong duality between the primal program and its dual program.

It should also be pointed out that we considered fairly general CTMDPs. Indeed, the state space was allowed to be arbitrary Polish, while the action space was Borel. The transition rates could be unbounded, and the cost rates may be unbounded from both above and below. Lastly, the class of history-dependent policies was taken into account.

Acknowledgements

This research was partially supported by the NSFC and GDUPS. We also thank the anonymous referee for constructive comments.

References

[1] ALIPRANTIS, C. AND BORDER, K. (2007). *Infinite Dimensional Analysis*. Springer, New York.
[2] ALTMAN, E. (1999). *Constrained Markov Decision Processes*. Chapman and Hall/CRC, Boca Raton.
[3] BERTSEKAS, D. P. AND SHREVE, S. E. (1978). *Stochastic Optimal Control*. Academic Press, New York.
[4] BERTSEKAS, D., NEDÍC, A. AND OZDAGLAR, A. (2003). *Convex Analysis and Optimization*. Athena Scientific, Belmont, MA.
[5] BOGACHEV, V. I. (2007). *Measure Theory*, Vol. I. Springer, Berlin.
[6] BOGACHEV, V. I. (2007). *Measure Theory*, Vol. II. Springer, Berlin.
[7] CLANCY, D. AND PIUNOVSKIY, A. B. (2005). An explicit optimal isolation policy for a determinisitc epidemic model. *Appl. Math. Comput.* **163,** 1109–1121.
[8] DUBINS, L. E. (1962). On extreme points of convex sets. *J. Math. Anal. Appl.* **5,** 237–244.
[9] FEINBERG, E. A. AND FEI, J. (2009). An inequality for variances of the discounted rewards. *J. Appl. Prob.* **46,** 1209–1212.
[10] FEINBERG, E. A. AND ROTHBLUM, U. G. (2012). Splitting randomized stationary policies in total-reward Markov decision processes. *Math. Operat. Res.* **37,** 129–153.
[11] GLEISSNER, W. (1988). The spread of epidemics. *Appl. Math. Comput.* **27,** 167–171.
[12] GUO, X. (2007). Constrained optimization for average cost continuous-time Markov decision processes. *IEEE Trans. Automatic Control* **52,** 1139–1143.
[13] GUO, X. AND HERNÁNDEZ-LERMA, O. (2009). *Continuous-time Markov Decision Processes*. Springer, Berlin.
[14] GUO, X. AND RIEDER, U. (2006). Average optimality for continuous-time Markov decision processes in Polish spaces. *Ann. Appl. Prob.* **16,** 730–756.
[15] GUO, X. AND SONG, X. (2011). Discounted continuous-time constrained Markov decision processes in Polish spaces. *Ann. Appl. Prob.* **21,** 2016–2049.
[16] GUO, X. AND ZHANG, L. (2011). Total reward criteria for unconstrained/constrained continuous-time Markov decision processes. *J. Systems Sci. Complex.* **24,** 491–505.
[17] GUO, X., HUANG, Y. AND SONG, X. (2012). Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent policies. *SIAM J. Control Optimization* **50,** 23–47.
[18] HERNÁNDEZ-LERMA, O. AND LASSERRE, J. B. (1996). *Discrete-time Markov Control Processes*. Springer, New York.
[19] HERNÁNDEZ-LERMA, O. AND LASSERRE, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.

[20] Hernández-Lerma, O. and Lasserre, J. B. (2000). Fatou's lemma and Lebesgue's convergence theorem for measures. *J. Appl. Math. Stoch. Anal.* **13,** 137–146.
[21] Himmelberg, C. J. (1975). Measurable relations. *Fund. Math.* **87,** 53–72.
[22] Himmelberg, C. J., Parthasarathy, T. and Van Vleck, F. S. (1976). Optimal plans for dynamic programming problems. *Math. Operat. Res.* **1,** 390–394.
[23] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nykodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitsth.* **31,** 235–253.
[24] Kitaev, M. (1986). Semi-Markov and jump Markov controlled models: average cost criterion. *Theory. Prob. Appl.* **30,** 272–288.
[25] Kitaev, M. and Rykov, V. V. (1995). *Controlled Queueing Systems*. CRC Press, Boca Raton, FL.
[26] Piunovskiy, A. B. (1997). *Optimal Control of Random Sequences in Problems with Constraints*. Kluwer, Dordrecht.
[27] Piunovskiy, A. B. (1998). A controlled jump discounted model with constraints. *Theory Prob. Appl.* **42,** 51–71.
[28] Piunovskiy, A. B. (2004). Optimal interventions in countable jump Markov processes. *Math. Operat. Res.* **29,** 289–308.
[29] Piunovskiy, A. and Zhang, Y. (2011). Accuracy of fluid approximation to controlled birth-and-death processes: absorbing case. *Math. Meth. Operat. Res.* **73,** 159–187.
[30] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time Markov decision processes with unbounded rates: the dynamic programming approach. Preprint. Available at http://arxiv.org/abs/1103.0134v1.
[31] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optimization* **49,** 2032–2061.
[32] Piunovskiy, A. and Zhang, Y. (2012). The transformation method for continuous-time Markov decision processes. *J. Optimization Theory Appl.* **154,** 691–712.
[33] Pliska, S. R. (1975). Controlled jump processes. *Stoch. Process Appl.* **3,** 259–282.
[34] Prieto-Rumeau, T. and Hernández-Lerma, O. (2008). Ergodic control of continuous-time Markov chains with pathwise constraints. *SIAM J. Control Optimization* **47,** 1888–1908.
[35] Rockafellar, R. T. (1974). *Conjugate Duality and Optimization*. SIAM, Philadelphia, PA.
[36] Varadarajan, V. S. (1958). Weak convergence of measures on separable metric spaces. *Sankhyā* **19,** 15–22.
[37] Yeh, J. (2006). *Real analysis: Theory of Measure and Integration*, 2nd edn. World Scientific, Hackensack, NJ.
[38] Zhang, Y. (2011). Convex analytic approach to constrained discounted Markov decision processes with non-constant discount factor. *TOP*, 31pp.
[39] Zhu, Q. (2008). Average optimality for continuous-time jump Markov decision processes with a policy iteration approach. *J. Math. Anal. Appl.* **339,** 691–704.
[40] Zhu, Q. and Prieto-Rumeau, T. (2008). Bias and overtaking optimality for continuous-time jump Markov decision processes in Polish spaces. *J. Appl. Prob.* **45,** 417–429.