Adv. Appl. Prob. **50,** 783–804 (2018)
*doi:*10.1017/apr.2018.36
© *Applied Probability Trust* 2018

RISK-SENSITIVE SEMI-MARKOV DECISION PROCESSES WITH GENERAL UTILITIES AND MULTIPLE CRITERIA

YONGHUI HUANG,* ** *Sun Yat-Sen University*
ZHAOTONG LIAN,*** *University of Macau*
XIANPING GUO,* **** *Sun Yat-Sen University*

Abstract

In this paper we investigate risk-sensitive semi-Markov decision processes with a Borel state space, unbounded cost rates, and general utility functions. The performance criteria are several expected utilities of the total cost in a finite horizon. Our analysis is based on a type of finite-horizon occupation measure. We express the distribution of the finite-horizon cost in terms of the occupation measure for each policy, wherein the discount is not needed. For unconstrained and constrained problems, we establish the existence and computation of optimal policies. In particular, we develop a linear program and its dual program for the constrained problem and, moreover, establish the strong duality between the two programs. Finally, we provide two special cases of our results, one of which concerns the discrete-time model, and the other the chance-constrained problem.

Keywords: Semi-Markov decision process; finite-horizon cost; occupation measure; expected utility; linear program

2010 Mathematics Subject Classification: Primary 90C40
Secondary 93E20

1. Introduction

As is well known, risk-sensitive control is a subject of significant interest in the field of Markov decision processes (MDPs) and has received much attention; see, for example, [2], [5], [8], and [14] for average cost discrete-time MDPs (DTMDPs), [2], [7], and [13] for infinite-horizon discounted cost DTMDPs, [2] and [7] for finite-horizon discounted or undiscounted cost DTMDPs, [10] and [22] for average cost continuous-time MDPs (CTMDPs), and [10] for finite-horizon cost and infinite-horizon discounted cost CTMDPs. Regarding risk-sensitive semi-Markov decision processes (SMDPs), to the best of the authors' knowledge this issue was only addressed in [6], in which the authors concentrated on a long-run average cost. In this paper we also study risk-sensitive SMDPs, but we focus on a finite-horizon total cost. We assume that the state space is a Borel space, the cost rate function is unbounded from above and from below, and the utility functions are general convex or concave functions. The performance criteria are several expected utilities of a finite-horizon total cost. We consider both unconstrained and constrained problems.

Received 17 May 2017; revision received 30 March 2018.

* Postal address: School of Mathematics, Sun Yat-Sen University, Guangzhou, 510275, China.
** Email address: hyongh5@mail.sysu.edu.cn
*** Postal address: Faculty of Business Administration, University of Macau, Macau, China.
Email address: lianzt@umac.mo
**** Email address: mcsgxp@mail.sysu.edu.cn

We adopt the convex analytic approach to solve our problems. Our work is related to that of Haskell and Jain [13]. In fact, the authors investigated risk-aware DTMDPs with general risk functions. They considered the problem of minimizing the risk of infinite-horizon discounted cost, but they focused on minimizing the risk of finite-horizon discounted cost with a sufficiently large time horizon instead. Whereas other authors employed a dynamic programming method commonly used for risk-sensitive MDPs and SMDPs (see [2], [5]–[8], [10], [14], and [22]), Haskell and Jain [13] developed a convex analytic approach to solve their problem from a fresh viewpoint: they augmented the state space, expressed the distribution of the finite-horizon total cost in terms of the infinite-horizon discounted occupation measure for each policy, and formulated the problem as an infinite-dimensional linear program in occupation measures. Compared with their work, however, our work has the following features.

- Our model is a continuous-time SMDP, which is more general than the model of DTMDPs in [13]. As is known, SMDPs allow the sojourn times to be arbitrarily distributed and, hence, SMDPs are well suited to model a variety of systems such as production scheduling, reliability, and maintenance.

- The cost rate function can be unbounded from above and below, whereas the one-stage cost function of [13] is assumed to be nonnegative and bounded from above. Thus, some new conditions and a framework different from those of [13] are forthcoming in our setup. We remark that the cases of unbounded costs arise in many real-world situations such as inventory systems (see [15, p. 69]), and have been widely studied in a risk-neutral context.

- We consider a finite-horizon total undiscounted cost, whereas an infinite-horizon total discounted cost was treated in [13]. Note that, for continuous-time models, finite-horizon problems are usually more complex and receive less attention than infinite-horizon problems due to the fact that the optimal policies for finite-horizon problems are nonstationary (depending on the time horizon). On the other hand, finite-horizon optimality is of particular interest in various applications since the lifetime of most systems in reality is usually finite and, moreover, it is often used to approximate the infinite-horizon optimality, as carried out in [13].

In keeping with the convex analytic approach proposed in [13] for risk-aware DTMDPs, we express the distribution of the finite-horizon total cost in terms of the occupation measure under each policy. In [13, Theorem 3.5], the distribution of the finite-horizon total cost was expressed in terms of the infinite-horizon discounted occupation measure, wherein the discount factor was indispensable. To adjust the approach to suit our undiscounted case, we introduce a type of finite-horizon undiscounted occupation measure instead and then for each policy, express the distribution of the finite-horizon total cost in terms of this type of occupation measure, wherein the discount factor is not needed. It is worth mentioning that our approach is also suitable for the discounted case.

Based on the treatment above, we write the expected utilities of the finite-horizon total cost in terms of the occupation measure for each policy under some finiteness conditions and further represent the expected utilities as the expected random-horizon total cost with new cost functions. For the unconstrained problem, we establish the Bellman equation and the existence of optimal policies under some compactness-continuity conditions. We also propose a method for the computation of optimal policies. For the constrained problem, we prove the compactness of the set of occupation measures and the continuity of the expected utilities in occupation measures with respect to w-weak topology under another set of

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

compactness-continuity conditions, which leads to the existence of constrained-optimal policies. Moreover, we develop a linear programming formulation together with its dual program for the constrained problem. We further establish the strong duality between the prime and the dual programs. Finally, two interesting special cases of our results are provided. One is on the discrete-time case, which shows that our analysis is distinct from the previous one. Another concerns the chance-constrained problem, in which the probabilities that the total cost exceeds some particular levels are constrained.

The rest of the paper is organized as follows. In Section 2 we formulate the risk-sensitive control model and propose our problems. In Section 3 we provide some preliminary analysis. Section 4 contains the solution to the unconstrained problem, whereas Section 5 contains the solution to the constrained problem. In Section 6 we present two special cases of our results. We conclude with Section 7.

2. The control model

Notation. If X is a Borel space, we denote by $\mathcal{B}(X)$ the Borel σ-algebra, by $\mathcal{P}(X)$ the set of all probability measures on $\mathcal{B}(X)$, and by $\mathbf{1}_D(x)$ the indicator function on the set $D \subseteq X$. Moreover, let $\mathbb{R} = (-\infty, +\infty)$ and $\mathbb{R}_+ = [0, +\infty)$.

A risk-sensitive SMDP is a model with a collection

$$\{E, A, (A(x),\ x \in E), Q(\cdot, \cdot \mid x, a), c(s, x, a), u_0(\cdot)\},$$

where E is the state space and A is the action set, which are assumed to be Borel spaces, respectively; $A(x) \in \mathcal{B}(A)$ denotes the set of admissible actions at state $x \in E$; $Q(\cdot, \cdot \mid x, a)$ is the semi-Markov kernel on $\mathbb{R}_+ \times E$ given K, where $K = \{(x, a) \mid x \in E, a \in A(x)\}$ denotes the set of admissible state-action pairs, assuming that $K \in \mathcal{B}(E \times A)$ and there exists a measurable mapping $\phi\colon \mathbb{R}_+ \times E \to A$ such that $(x, \phi(t, x)) \in K$; the function $c(s, x, a)$ on $\mathbb{R}_+ \times K$ represents the cost/loss rate at time s, depending on the state and action. We allow the cost/loss function c to be unbounded from below and from above. The function $u_0(\cdot)$ on $\mathbb{R}$ denotes the central utility function.

If, furthermore, a collection $\{u_i(\cdot), d_i,\ i = 1, \dots, q\}$ is taken into account, where $u_1, \dots, u_q$ represent the constrained utility functions on $\mathbb{R}$ and d_i are constraint constants, the model is referred to as a constrained risk-sensitive SMDP. As usual, the utility functions are assumed to be continuous, strictly increasing, and convex or concave; see [1, Section 3.4] and [2] and the references therein.

Remark 2.1. The utility-constrained problems arise when we wish to minimize/maximize the total cost/reward under more than one utility function. For example, for a corporation, the central utility is usually the economic benefits, while the constrained utilities that the corporation have to consider are environment benefits, society benefits, employee welfare, and so on. The utility functions are chosen individually and should to some extent also reflect his/her risk attitude and the degree of risk preference; see Remark 2.3 for further details.

The evolution of an SMDP is as follows. Initially, the system occupies some state $x_0 \in E$ at some time $t_0 \in \mathbb{R}_+$, and the controller chooses an action $a_0 \in A(x_0)$ according to some rule. As a consequence, the system jumps to state $x_1 \in E$ after a sojourn time $\theta_1 \in \mathbb{R}_+$ in x_0, in which the transition law is subject to the semi-Markov kernel Q. At time $(t_0 + \theta_1)$, the controller chooses an action $a_1 \in A(x_1)$ according to some rule and the same sequence of events occurs. The SMDP evolves in this way, and we obtain a history $h_n = (t_0, x_0, a_0, \theta_1, x_1, a_1, \dots, \theta_n, x_n)$

of the SMDP up to the nth jump time. Equivalently, a history h_n can be written as $h_n = (t_0, x_0, a_0, t_1, x_1, a_1, \dots, t_n, x_n)$, where $t_k = t_{k-1} + \theta_k$, $k = 1, 2, \dots, n$, representing the jump times of the SMDP. For each $n \geq 0$, let $H_n = (\mathbb{R}_+ \times E \times A)^n \times (\mathbb{R}_+ \times E)$ denote the set of all histories of the SMDP up to the nth jump, which is endowed with the Borel σ-algebra.

To regulate the action-selection, we will introduce policies. A policy $\pi = \{\pi_n,\ n \geq 0\}$ is a sequence of stochastic kernels π_n on A given H_n satisfying $\pi_n(A(x_n) \mid h_n) = 1$ for every $h_n \in H_n$, $n = 0, 1, \dots$. The set of all policies is denoted by Π. Below we state some special policies.

Definition 2.1. A randomized Markov policy $\pi = \{\phi_n\}$ is a sequence of stochastic kernels ϕ_n on A given $\mathbb{R}_+ \times E$ such that $\phi_n(A(x_n) \mid t_n, x_n) = 1$ for every $(t_n, x_n) \in \mathbb{R}_+ \times E$ and $n \geq 0$. If ϕ_n are all independent of n, it is said to be randomized jump-stationary. In this case, we write $\pi = \{\phi, \phi, \dots\}$ as ϕ for simplicity. If, furthermore, $\phi(\cdot \mid t, x)$ is a Dirac measure for each (t, x), it is said to be deterministic jump-stationary, and we write $\phi(\cdot \mid t, x)$ as $\phi(t, x)$.

We denote by Π^{RS} and Π^{DS} the families of all randomized jump-stationary and deterministic jump-stationary policies, respectively.

Let $\Omega = (\mathbb{R}_+ \times E \times A)^\infty$ be a sample space and $\mathcal{F} = \mathcal{B}(\Omega)$ be the Borel σ-algebra. On the measurable space $(\Omega, \mathcal{F})$, we define a sequence of random variables (RVs) $\{T_n, \Theta_{n+1}, X_n, A_n, n \geq 0\}$ by

$$T_0(\omega) = t_0, \qquad T_{n+1}(\omega) = t_0 + \theta_1 + \cdots + \theta_{n+1}, \qquad \Theta_{n+1}(\omega) = \theta_{n+1},$$
$$X_n(\omega) = x_n, \qquad A_n(\omega) = a_n$$

for each $n \geq 0$ and any trajectory $\omega = (t_0, x_0, a_0, \theta_1, x_1, a_1, \dots, \theta_n, x_n, a_n, \dots) \in \Omega$. Here, T_n denotes the RV of the nth jump time of the SMDP; Θ_{n+1}, the RV of the sojourn time between the nth and the $(n+1)$th jumps; X_n, the RV of the post-jump state at T_n; and A_n, the RV of the action chosen at T_n. Now, given the semi-Markov kernel Q, an initial time-state distribution $\mu \in \mathcal{P}(\mathbb{R}_+ \times E)$, and a policy $\pi \in \Pi$, by the Ionescu Tulcea theorem (see [1, Proposition B.2.5], there exists a unique probability measure $\mathbb{P}^\pi_\mu$ on $(\Omega, \mathcal{F})$ such that

$$\begin{aligned}\mathbb{P}^\pi_\mu(T_0 \in \mathrm{d}t, X_0 \in \mathrm{d}x) &= \mu(\mathrm{d}t,\ \mathrm{d}x),\\ \mathbb{P}^\pi_\mu(A_n \in \Gamma \mid T_0, X_0, A_0, \dots, T_n, X_n) &= \pi_n(\Gamma \mid T_0, X_0, A_0, \dots, T_n, X_n),\\ \mathbb{P}^\pi_\mu(\Theta_{n+1} \leq s, X_{n+1} \in B \mid T_0, X_0, A_0, \dots, T_n, X_n, A_n) &= Q(s, B \mid X_n, A_n) \qquad (2.1)\end{aligned}$$

for each $s \in \mathbb{R}_+$, $B \in \mathcal{B}(E)$, $\Gamma \in \mathcal{B}(A)$, and $n \geq 0$. The expectation operator with respect to $\mathbb{P}^\pi_\mu$ is denoted by $\mathbb{E}^\pi_\mu$. When μ is a Dirac measure $\delta_{\{(t,x)\}}(\cdot)$, i.e. $\mu(\{(t, x)\}) = 1$, we use the notation $\mathbb{P}^\pi_{(t,x)}$ and $\mathbb{E}^\pi_{(t,x)}$ instead of $\mathbb{P}^\pi_\mu$ and $\mathbb{E}^\pi_\mu$, respectively.

Let $N(t) := \sum_{n\geq 1} \mathbf{1}_{\{T_n \leq t\}}$ represent the number of jumps of the SMDP until time t, and let $T_\infty := \lim_{n\to\infty} T_n$ be the explosive time of the SMDP. Note that T_∞ may be finite. We do not intend to consider the controlled process after the moment T_∞. For each $t < T_\infty$, let $\xi_t(\omega) = \sum_{n\geq 0} \mathbf{1}_{\{T_n \leq t < T_{n+1}\}}(\omega) X_n(\omega)$ and $\vartheta_t(\omega) = \sum_{n\geq 0} \mathbf{1}_{\{T_n \leq t < T_{n+1}\}}(\omega) A_n(\omega)$ denote the underlying continuous-time state and action processes, respectively. Hereafter, we consider a T-horizon SMDP (with $T > 0$) and, naturally, assume a finite number of jumps until time T. Thus, we propose Assumption 2.1.

Assumption 2.1. *There exist constants $\delta > 0$ and $\varepsilon > 0$ such that*

$$\sup_{(x,a)\in K} Q(\delta, E \mid x, a) \leq 1 - \varepsilon.$$

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

Remark 2.2. (i) Assumption 2.1 is widely employed in SMDPs as a regular condition; see, for example, [19] for the expected infinite-horizon discounted reward criteria, [3] and [19] for the expected long-run average reward criteria, [16] for a probability criterion, and [17] for an expected finite-horizon reward criterion, and the references therein.

(ii) An equivalent condition to Assumption 2.1 is that $\sup_{(x,a)\in K} Q(0, E \mid x, a) < 1$, which, however, is easy to verify in various applications. For example, in a DTMDP, $Q(0, E \mid x, a) \equiv 0$ and, thus, Assumption 2.1 is fulfilled.

Assumption 2.1 means that the sojourn time at any state and action exceeds δ with a probability at least ε and, thus, there cannot be an infinite number of jumps over any finite time-horizon with probability 1. Formally, following the proof of [21, Proposition 3.2.2], Assumption 2.1 implies that

$$\mathbb{E}_{\mu}^{\pi}[N(T)] \leq \frac{T/\delta + 1}{\varepsilon} < \infty \quad \text{for all } \pi \in \Pi. \tag{2.2}$$

Therefore, we have $\mathbb{P}_{\mu}^{\pi}(\{N(T) < \infty\}) = 1$ for all $\pi \in \Pi$ and, hence, the T-horizon total cost $\int_{T_0}^{T} c(s, \xi_s, \vartheta_s)\, \mathrm{d}s$ is well defined. In this context, for each initial distribution $\mu \in \mathcal{P}([0, T] \times E)$ and policy $\pi \in \Pi$, we can now define the expected utilities of T-horizon cost by

$$J_i^{\pi}(\mu) := \mathbb{E}_{\mu}^{\pi}\bigg[u_i\bigg(\int_{T_0}^{T} c(s, \xi_s, \vartheta_s)\, \mathrm{d}s\bigg)\bigg], \qquad i = 0, \dots, q,$$

whenever the expectations are well defined. In particular, if $\mu(\{(t, x)\}) = 1$ for some $(t, x) \in [0, T] \times E$, we write $J_i^{\pi}(\mu)$ as $J_i^{\pi}(t, x)$.

Remark 2.3. (i) Note that the expected utility incorporates the mean and variance of the total cost and, thus, characterizes the risk preference. To see this, let Y represent some type of cost and u be a utility function. Using the Taylor rule, we can write the expected utility of Y as

$$\mathbb{E}[u(Y)] \approx u(\mathbb{E}[Y]) + \tfrac{1}{2}u''(\mathbb{E}[Y])\, \mathrm{var}[Y].$$

The cases of convex or concave u describe the risk-averse or risk-seeking preference. For details, see [1, pp. 70–72] or [2, pp. 3, 4].

(ii) The exponential utility $u(y) = (1/\gamma)\mathrm{e}^{\gamma y}(\gamma \neq 0)$ is one of the most popular utility functions, and has received much attention in the literature of risk-sensitive MDPs; see [2], [5], [7], [8], and [14]. Other popular utility functions include the logarithmic utility $u(y) = \log(\gamma y)(\gamma > 0)$, the power utility $u(y) = (1/\gamma)y^{\gamma}(\gamma \neq 0)$, and the quadratic utility $u(y) = y + \gamma y^2(\gamma \neq 0)$. Note that the sign of the coefficient γ reflects the risk preference of an individual, while the absolute value of γ determines the level of risk tolerance; see [1, pp. 70–72] for details.

We are now ready to state our unconstrained and constrained risk-sensitive problems. The unconstrained risk-sensitive problem is of the form

$$\min_{\pi\in\Pi} J_0^{\pi}(\mu), \tag{2.3}$$

whereas the constrained risk-sensitive problem is as follows:

$$\begin{aligned} \min_{\pi\in\Pi} \quad & J_0^{\pi}(\mu) \\ \text{subject to} \quad & J_i^{\pi}(\mu) \leq d_i, \qquad i = 1, \dots, q. \end{aligned} \tag{2.4}$$

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

In what follows, we first carry out some preliminary analysis in Section 2, and then provide solutions to the unconstrained problem and the constrained problem in Sections 4 and 5, respectively.

3. Preliminary analysis

In this section, our aim is threefold:

- introduce additional variables that summarize the accumulated past costs at jump times;
- introduce a type of finite-horizon occupation measure;
- express the distribution of the finite-horizon total cost in terms of occupation measure under each policy.

3.1. State augmentation

As is well known, problems of minimizing the expected utility of a total cost can be handled by state augmentation through introducing an auxiliary state variable that keeps track of the cumulative past cost; see [2] and [13]. The reason is that an optimal policy for problems of this type usually depends on the accumulated past cost. In our setup of finite-horizon SMDPs, we will also introduce such variables. Let Λ_0 denote the initial cost, which is an RV (although it is determinate in most cases). For every $n \geq 1$, we define the cumulative costs Λ_n until the nth jump before the terminal time T by

$$\Lambda_n := \Lambda_0 + \sum_{k=0}^{n-1} \int_{T\wedge T_k}^{T\wedge T_k+\Theta_{k+1}\wedge(T-T\wedge T_k)} c(s, X_k, A_k)\,\mathrm{d}s.$$

Since the present costs will not accumulate when time T is reached, we have

$$\Lambda_{N(T)+1} = \Lambda_0 + \int_{T_0}^{T} c(s, \xi_s, \vartheta_s)\,\mathrm{d}s.$$

Now we augment the state space with two new state variables, namely T_n and Λ_n, to keep track of the jump times and accumulated past costs, respectively. The transition law of the augmented state process $\{T_n, \Lambda_n, X_n,\ n = 0, 1, \dots, N(T)\}$ is governed by

$$\mathbb{Q}(B \times C \times D \mid t, \lambda, x, a) := \int_0^{T-t} \mathbf{1}_B(t+s)\delta_{\{\lambda+\int_t^{t+s} c(z,x,a)\,\mathrm{d}z\}}(C)Q(\mathrm{d}s, D \mid x, a)$$

for every $B \times C \times D \in \mathcal{B}([0, T] \times \mathbb{R} \times E)$ and $(t, \lambda, x, a) \in \mathbb{K} := [0, T] \times \mathbb{R} \times K$. Note that the transition kernel $\mathbb{Q}$ is substochastic since $\mathbb{Q}([0, T] \times \mathbb{R} \times E \mid t, \lambda, x, a) = Q(T - t, E \mid x, a) \leq 1$. If we introduce an isolated state Δ along with an isolated action a_Δ such that $\mathbb{Q}(\Delta \mid t, \lambda, x, a) = 1 - \mathbb{Q}([0, T]\times\mathbb{R}\times E \mid t, \lambda, x, a)$ for $(t, \lambda, x, a) \in \mathbb{K}$, and $\mathbb{Q}(\Delta \mid \Delta, a_\Delta) = 1$, $\mathbb{Q}$ becomes a stochastic kernel for the extended state space $[0, T] \times \mathbb{R} \times E \cup \{\Delta\}$ and the action space $A \cup \{a_\Delta\}$. The policies are also extended to include the information on Λ_n since the Λ_n summarize the useful history information for the controllers. For simplicity, we will still use the same policy notation to denote the generalized policies.

In a more general framework, given an initial time-cost state distribution $\nu \in \mathcal{P}([0, T] \times \mathbb{R} \times E)$ and a policy $\pi \in \Pi$, by the Ionescu Tulcea theorem [1, Proposition B.2.5] again, there exist a discrete-time process $\{T'_n, \Lambda'_n, X'_n, A'_n,\ n = 0, 1, \dots\}$ and a corresponding probability

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

space $(\Omega', \mathcal{F}', \mathbb{P}^{\pi}_{\nu})$ associated with the extended kernel $\mathbb{Q}$. Clearly, the first entrance time of the process $\{T'_n, \Lambda'_n, X'_n\}$ into Δ is $N(T)+1$ and, thus, $(T'_n, \Lambda'_n, X'_n, A'_n)$ coincides with $(T_n, \Lambda_n, X_n, A_n)$ for every $n \le N(T)$, while $(T'_n, \Lambda'_n, X'_n) = \Delta$ and $A'_n = a_{\Delta}$ for all $n \ge N(T)+1$. The expected utilities of the finite-horizon total cost are now defined by

$$\mathbb{J}^{\pi}_i(\nu) := \mathbb{E}^{\pi}_{\nu}[u_i(\Lambda_{N(T)+1})], \qquad \pi \in \Pi,\ i = 0, \dots, q,$$

provided that the expectations are well defined. Accordingly, the unconstrained problem (2.3) is extended to a more general form as

$$\min_{\pi \in \Pi} \mathbb{J}^{\pi}_0(\nu), \tag{3.1}$$

while the constrained problem (2.4) is extended to

$$\begin{aligned} &\min_{\pi \in \Pi} && \mathbb{J}^{\pi}_0(\nu) \\ &\text{subject to} && \mathbb{J}^{\pi}_i(\nu) \le d_i, \qquad i = 1, \dots, q, \end{aligned} \tag{3.2}$$

3.2. Occupation measures

Next we introduce a type of finite-horizon occupation measure and establish a one-to-one correspondence between randomized jump-stationary policies and occupation measures.

Definition 3.1. The occupation measure $\eta^{\pi}_{\nu}(\cdot)$ of a policy π is a nonnegative measure on $\mathcal{B}([0, T] \times \mathbb{R} \times E \times A)$ concentrated on $\mathbb{K}$, which is defined by

$$\eta^{\pi}_{\nu}(B, C, D, \Gamma) := \mathbb{E}^{\pi}_{\nu}\Bigg[\sum_{n=0}^{N(T)} \mathbf{1}_{\{T_n \in B,\ \Lambda_n \in C,\ X_n \in D,\ A_n \in \Gamma\}}\Bigg], \tag{3.3}$$

where B, C, D, and Γ are measurable subsets of $[0, T]$, $\mathbb{R}$, E, and A, respectively.

We denote by $\mathcal{D} := \{\eta^{\pi}_{\nu} : \pi \in \Pi\}$ the space of all occupation measures. Note that, by (2.2), the occupation measures η^{π}_{ν} are bounded in π under Assumption 2.1.

Remark 3.1. (i) The occupation measures are in fact state-action frequencies. They are widely used in MDPs so as to transform a stochastic dynamic control problem to a static optimization problem; see [9], [11], [13], [15], [18], and [19].

(ii) The occupation measures defined above are in fact finite-horizon undiscounted occupation measures, which are different from the infinite-horizon discounted occupation measures used to analyze the risk-aware DTMDPs in [13]; see Section 6 for further details.

In the next theorem we state some characterizations of the elements of $\mathcal{D}$.

Theorem 3.1. *Under Assumption 2.1, the following assertions hold.*

(i) *For a fixed policy $\pi \in \Pi$, η^{π}_{ν} is a finite measure satisfying*

$$\begin{aligned} \hat{\eta}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y) &= \nu(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y) \\ &\quad + \int_{[0,T]\times\mathbb{R}\times E\times A} \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a)\eta(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a), \end{aligned} \tag{3.4}$$

where $\hat{\eta}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y) := \eta(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y, A)$ is the marginal of η on $[0, T] \times \mathbb{R} \times E$.

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

(ii) *Conversely, if a finite measure* η *on* $\mathcal{B}([0,T]\times\mathbb{R}\times E\times A)$ *concentrated on* $\mathbb{K}$ *satisfies (3.4), there is a randomized jump-stationary policy* $\phi\in\Pi^{\mathrm{RS}}$ *such that* $\eta_\nu^\phi=\eta$ *and* ϕ *can be taken from the disintegration of* η *with respect to its marginal* $\hat\eta$*, i.e.*

$$\eta(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a)=\phi(\mathrm{d}a\mid t,\lambda,x)\hat\eta(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x). \tag{3.5}$$

Moreover, if $\varphi\in\Pi^{\mathrm{RS}}$ *is such that* $\eta_\nu^\varphi=\eta$ *then* φ *is a version of* ϕ *above with respect to* $\hat\eta$*, and*

$$\eta_\nu^\varphi(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a)=\varphi(\mathrm{d}a\mid t,\lambda,x)\hat\eta_\nu^\varphi(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x). \tag{3.6}$$

Proof. Under Assumption 2.1, (2.2) indicates that $\{T_n',\Lambda_n',X_n',A_n',\ n=0,1,\ldots\}$ is a DTMDP absorbing to Δ. Thus, the theory for absorbing MDPs can be applied in our setup. For instance, (i) and the first assertion of (ii) follow from [15, Lemmas 4.2 and 4.3].

To prove the second assertion of (ii), suppose that we have $D\in\mathcal{B}([0,T]\times\mathbb{R}\times E)$ such that $\hat\eta(D)>0$ and $\varphi(\Gamma\mid t,\lambda,x)\neq\phi(\Gamma\mid t,\lambda,x)$ for every $(t,\lambda,x)\in D$ and some $\Gamma\in\mathcal{B}(A)$. Since $\eta(\cdot)=\eta_\nu^\varphi(\cdot)$, we have $\hat\eta(\cdot)=\hat\eta_\nu^\varphi(\cdot)$. However, we find that

$$\eta_\nu^\varphi(D,\Gamma)=\int_D\varphi(\Gamma\mid t,\lambda,x)\hat\eta_\nu^\varphi(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x)\neq\int_D\phi(\Gamma\mid t,\lambda,x)\hat\eta(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x)=\eta(D,\Gamma),$$

which contradicts with $\eta(\cdot)=\eta_\nu^\varphi(\cdot)$. Thus, we must have, for every $\Gamma\in\mathcal{B}(A)$, $\varphi(\Gamma\mid t,\lambda,x)=\phi(\Gamma\mid t,\lambda,x)$, $\hat\eta$-almost every $(t,\lambda,x)\in[0,T]\times\mathbb{R}\times E$, i.e. φ is a version of ϕ. This fact together with (3.5) leads to (3.6). □

3.3. Distribution of the finite-horizon cost

We now express the distribution of the finite-horizon total cost $\Lambda_{N(T)+1}$ in terms of the occupation measure η_ν^π under each policy π.

Theorem 3.2. *For every* $\pi\in\Pi$ *and* $D\in\mathcal{B}(\mathbb{R})$*,*

$$\begin{aligned}&\mathbb{P}_\nu^\pi(\Lambda_{N(T)+1}\in D)\\&\quad=\int_{[0,T]\times\mathbb{R}\times E\times A}[(1-Q(T-s,E\mid y,a))\,\mathbf{1}_{\{\gamma+\int_s^T c(z,y,a)\,\mathrm{d}z\in D\}}]\eta_\nu^\pi(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y,\,\mathrm{d}a).\end{aligned}$$

Proof. For every $\pi\in\Pi$, we have

$$\begin{aligned}&\mathbb{P}_\nu^\pi(\Lambda_{N(T)+1}\in D)\\&\quad=\sum_{n=0}^\infty\mathbb{P}_\nu^\pi(\Lambda_{n+1}\in D,N(T)=n)\\&\quad=\sum_{n=0}^\infty\mathbb{P}_\nu^\pi\bigg(\Lambda_n+\int_{T_n}^T c(s,X_n,A_n)\,\mathrm{d}s\in D,T_n\le T<T_{n+1}\bigg)\\&\quad=\sum_{n=0}^\infty\mathbb{E}_\nu^\pi\bigg[\mathbb{P}_\nu^\pi\bigg(\Lambda_n+\int_{T_n}^T c(s,X_n,A_n)\,\mathrm{d}s\in D,T_n\le T<T_{n+1}\mid T_n,\Lambda_n,X_n,A_n\bigg)\bigg]\\&\quad=\sum_{n=0}^\infty\mathbb{E}_\nu^\pi[\mathbf{1}_{\{T_n\le T\}}\mathbf{1}_{\{\Lambda_n+\int_{T_n}^T c(s,X_n,A_n)\,\mathrm{d}s\in D\}}\mathbb{P}_\nu^\pi(\Theta_{n+1}>T-T_n\mid T_n,\Lambda_n,X_n,A_n)]\end{aligned}$$

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \mathbb{E}_\nu^\pi [\mathbf{1}_{\{T_n \le T\}} \mathbf{1}_{\{\Lambda_n + \int_{T_n}^T c(s, X_n, A_n)\,\mathrm{d}s \in D\}} (1 - Q(T - T_n \mid X_n, A_n))] \\
&= \int_{[0,T]\times\mathbb{R}\times E\times A} [(1 - Q(T - s, E \mid y, a))\, \mathbf{1}_{\{\gamma + \int_s^T c(z,y,a)\,\mathrm{d}z \in D\}}] \eta_\nu^\pi(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y, \mathrm{d}a),
\end{aligned}
$$

where the first equality follows from Assumption 2.1, the fourth equality is due to the properties of conditional expectations, the fifth equality is from (2.1), and the last is due to the definition of occupation measures. The proof is complete. □

Remark 3.2. In fact, expressing the distribution of the total cost in terms of occupation measure under each policy plays a key role in analyzing our risk-sensitive problems. This idea is inspired by that of Haskell and Jain [13] for risk-aware DTMDPs. To make the approach of [13] for the discounted case fit our undiscounted case, our expression here is based on the finite-horizon undiscounted occupation measures.

According to Theorem 3.2, for every $i = 0, \dots, q$ and $\pi \in \Pi$, we can write $\mathbb{J}_i^\pi(\nu)$ in terms of the occupation measure η_ν^π as

$$
\begin{aligned}
\mathbb{J}_i^\pi(\nu) &= \int_{\mathbb{R}} u(\lambda) \int_{[0,T]\times\mathbb{R}\times E\times A} [(1 - Q(T - s, E \mid y, a))\, \mathbf{1}_{\{\gamma + \int_s^T c(z,y,a)\,\mathrm{d}z \in \mathrm{d}\lambda\}}] \\
&\qquad\qquad \times \eta_\nu^\pi(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y, \mathrm{d}a) \\
&= \int_{[0,T]\times\mathbb{R}\times E\times A} \tilde{c}_i(s, \gamma, y, a) \eta_\nu^\pi(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y, \mathrm{d}a), \qquad (3.7)
\end{aligned}
$$

where $\tilde{c}_i$ are defined by

$$
\tilde{c}_i(s, \gamma, y, a) := (1 - Q(T - s, E \mid y, a)) u_i\biggl(\gamma + \int_s^T c(z, y, a)\,\mathrm{d}z\biggr), (s, \gamma, y, a) \in \mathbb{K}.
$$

From (3.7), we see that $\mathbb{J}_i^\pi(\nu)$ is linear in the occupation measure η_ν^π for each $i = 0, \dots, q$, which implies that the optimization problems (3.1) and (3.2) may be transformed to linear program problems, as carried out in Section 5. Moreover, when we derive (3.7), we have exchanged the order of the double integration. To ensure that the exchange of the integration order is valid, by the Fubini theorem, the double integral should be finite, which, however, is not guaranteed at this moment. To achieve this, we propose the following conditions.

Assumption 3.1. *There exist a weight (or bounding) function* $w \ge 1$ *on* $[0, T] \times \mathbb{R} \times E$ *and two constants* $M_1, M_2 > 0$ *such that:*

(i) $|\tilde{c}_i(t, \lambda, x, a)| \le M_1 w(t, \lambda, x)$ *for all* $(t, \lambda, x, a) \in \mathbb{K}$*;*

(ii) $\int_{[0,T]\times\mathbb{R}\times E} w(t, \lambda, x) \nu(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x) < +\infty$*;*

(iii) $\mathbb{E}_{(t,\lambda,x)}^\pi [\sum_{n=0}^{N(T)} w(T_n, \Lambda_n, X_n)] \le M_2 w(t, \lambda, x)$ *for all* $\pi \in \Pi$.

Remark 3.3. (i) If the cost rate $c(x, a)$ and the initial cost Λ_0 are both bounded, the cost functions $\tilde{c}_i$ will be bounded. In this case, the weight function w can be taken as $w(\cdot) \equiv 1$ in Assumption 3.1.

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

(ii) A sufficient condition for Assumption 3.1(iii) is that there exists a positive constant $\rho < 1$ such that

$$\int_0^{T-t}\int_E w\biggl(t+s, \lambda + \int_t^{t+s} c(z, x, a)\,\mathrm{d}z, y\biggr)Q(\mathrm{d}s,\ \mathrm{d}y \mid x, a) \leq \rho w(t, \lambda, x)$$

for all $(t, \lambda, x, a) \in \mathbb{K}$, under which we have

$$\mathbb{E}^{\pi}_{(t,\lambda,x)}\Biggl[\sum_{n=0}^{N(T)} w(T_n, \Lambda_n, X_n)\Biggr] \leq \frac{w(t, \lambda, x)}{1-\rho}.$$

In particular, the constant ρ is allowed to be greater than 1 when the SMDP models reduce to the DTMDP models.

(iii) Assumptions 3.1(ii) and 3.1(iii) imply that

$$\mathbb{E}^{\pi}_{\nu}[N(T)] \leq M_2 \int_{[0,T]\times\mathbb{R}\times E} w(t, \lambda, x)\nu(\mathrm{d}t,\ \mathrm{d}\lambda,\ \mathrm{d}x) < +\infty$$

and, thus, Assumption 2.1 is unnecessary whenever Assumption 3.1 is imposed.

Under Assumption 3.1, we see that for every $\pi \in \Pi$,

$$\int_{\mathbb{K}} |\tilde{c}_i(s, \gamma, y, a)|\eta(\mathrm{d}s,\ \mathrm{d}\gamma,\ \mathrm{d}y,\ \mathrm{d}a) \leq M_1M_2 \int_{[0,T]\times\mathbb{R}\times E} w(t, \lambda, x)\nu(\mathrm{d}t,\ \mathrm{d}\lambda,\ \mathrm{d}x) < +\infty.$$

Thus, by the Fubini theorem, the equality of (3.7) is validated.

For convenience, we introduce

$$V^{\pi}(\tilde{c}_i, \nu) := \mathbb{E}^{\pi}_{\nu}\Biggl[\sum_{n=0}^{N(T)} \tilde{c}_i(T_n, \Lambda_n, X_n, A_n)\Biggr], \qquad \pi \in \Pi,\ i = 0, 1, \dots, q.$$

It is obvious that $V^{\pi}(\tilde{c}_i, \nu) = \int_{\mathbb{K}}\tilde{c}_i(s, \gamma, y, a)\eta(\mathrm{d}s,\ \mathrm{d}\gamma,\ \mathrm{d}y,\ \mathrm{d}a)$. Thus, by (3.7), $\mathbb{J}^{\pi}_i(\nu)$ have the equivalent expressions $\mathbb{J}^{\pi}_i(\nu) = V^{\pi}(\tilde{c}_i, \nu)$ for all $\pi \in \Pi$ and $i = 0, \dots, q$. This shows that we have transformed the expected utilities as the expected random-horizon costs that are standard risk-neutral criteria. In this way, we can write the unconstrained problem (3.1) as

$$\min_{\pi\in\Pi} \quad V^{\pi}(\tilde{c}_0, \nu), \tag{3.8}$$

and the constrained problem (3.2) as

$$\begin{aligned} &\min_{\pi\in\Pi} && V^{\pi}(\tilde{c}_0, \nu) \\ &\text{subject to} && V^{\pi}(\tilde{c}_i, \nu) \leq d_i, \qquad i = 1, \dots, q. \end{aligned} \tag{3.9}$$

In what follows, we aim to solve the two problems. For the unconstrained problem, our goal is to find an optimal policy $\pi^* \in \Pi$ such that $V^{\pi^*}(\tilde{c}_0, \nu) = V^*(\tilde{c}_0, \nu)$, where $V^*(\tilde{c}_0, \nu) := \inf_{\pi\in\Pi} V^{\pi}(\tilde{c}_0, \nu)$ is the optimal value. For the constrained problem, our aim is to search for a constrained-optimal policy $\pi^* \in \mathbb{U}$ such that $V^{\pi}(\tilde{c}_0, \nu) = \inf_{\pi\in\mathbb{U}} V^{\pi}(\tilde{c}_0, \nu)$, where $\mathbb{U} := \{\pi \in \Pi \mid V^{\pi}(\tilde{c}_i, \nu) \leq d_i,\ i = 1, \dots, q\}$ is the feasible-policy set. It is natural to assume that $\mathbb{U} \neq \varnothing$. A necessary and sufficient condition for $\mathbb{U}$ to be nonempty will be discussed in Remark 5.3 below.

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

4. Solution to the unconstrained problem

In this section we use the dynamic programming approach to solve the unconstrained problem. Although the unconstrained problem can be also solved via the convex analytic approach as in Section 5 for the constrained problem below, we use the dynamic programming approach here to derive relevant results since they will play a key role in establishing the strong duality in Subsection 5.2, and they help to reveal the relationship between the solution to the primal program and the one to the dual program; see Remark 5.5.

For each $\pi \in \Pi$, each initial state $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$, and $n = 1, 2, \ldots$, let

$$V_n^\pi(\tilde{c}_0, t, \lambda, x) := \mathbb{E}_{(t,\lambda,x)}^\pi \Bigg[\sum_{k=0}^{n-1} \mathbf{1}_{\{T_k \le T\}} \, \tilde{c}_0(T_k, \Lambda_k, X_k, A_k) \Bigg],$$

$$V^\pi(\tilde{c}_i, t, \lambda, x) := \mathbb{E}_{(t,\lambda,x)}^\pi \Bigg[\sum_{n=0}^{N(T)} \tilde{c}_0(T_n, \Lambda_n, X_n, A_n) \Bigg],$$

$V_n^*(\tilde{c}_0, t, \lambda, x) := \inf_{\pi\in\Pi} V_n^\pi(\tilde{c}_0, t, \lambda, x)$ and $V^*(\tilde{c}_0, t, \lambda, x) := \inf_{\pi\in\Pi} V^\pi(\tilde{c}_0, t, \lambda, x)$ be the value functions. Moreover, for the weight function w from Assumption 3.1, let $\mathbb{B}_w(\mathbb{K})$ and $\mathbb{B}_w([0, T] \times \mathbb{R} \times E)$ denote the spaces of measurable functions v, respectively, with a finite w-norm, i.e. v satisfies

$$\sup_{(t,\lambda,x,a)\in\mathbb{K}} \frac{|v(t, \lambda, x, a)|}{w(t, \lambda, x)} < +\infty \quad \text{or} \quad \sup_{(t,\lambda,x)\in[0,T]\times\mathbb{R}\times E} \frac{|v(t, \lambda, x)|}{w(t, \lambda, x)} < +\infty,$$

respectively. Obviously, under Assumptions 3.1(i) and 3.1(iii), $V_n^\pi(\tilde{c}_0, \cdot)$ and $V^\pi(\tilde{c}_0, \cdot)$ are all in $\mathbb{B}_w([0, T] \times \mathbb{R} \times E)$, and $\lim_{n\to\infty} V_n^\pi(\tilde{c}_0, \cdot) = V^\pi(\tilde{c}_0, \cdot)$.

To solve the unconstrained problem, we need some compact-continuity conditions that are extensions of those for risk-neutral DTMDPs; see [15].

Assumption 4.1. (Compact-continuity condition.) *For each* $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$*:*

(i) $A(x)$ *is compact;*

(ii) $Q(t, E \mid x, a)$ *is upper semicontinuous in* $a \in A(x)$*;*

(iii) $c(t, x, a)$ *is lower semicontinuous in* $a \in A(x)$*;*

(iv) $\int_0^{T-t} \int_E v(t + s, \lambda + \int_t^{t+s} c(z, x, a)\,\mathrm{d}z, y) Q(\mathrm{d}s, \mathrm{d}y \mid x, a)$ *is lower semicontinuous in* $a \in A(x)$ *for any bounded measurable function* v *on* $[0, T] \times \mathbb{R} \times E$*;*

(v) $\int_0^{T-t} \int_E w(t + s, \lambda + \int_t^{t+s} c(z, x, a)\,\mathrm{d}z, y) Q(\mathrm{d}s, \mathrm{d}y \mid x, a)$ *is continuous in* $a \in A(x)$*.*

Remark 4.1. (i) Assumptions 4.1(ii) and 4.1(iii) along with the properties of u_i imply that all $\tilde{c}_i(t, \lambda, x, a)$ are lower semicontinuous in $a \in A(x)$ for each $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$.

(ii) As proved in [15, Lemma 8.3.7], we can similarly conclude that, under Assumptions 4.1(iv) and 4.1(v), $\int_{[0,T]\times\mathbb{R}\times E} v(s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a)$ is continuous in $a \in A(x)$ for each $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$ and $v \in \mathbb{B}_w([0, T] \times \mathbb{R} \times E)$.

We are now ready to state the main optimality results for the unconstrained problem.

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

Theorem 4.1. *Suppose that Assumptions 3.1(i), 3.1(iii), and 4.1 hold.*

(i) *For all* $n \ge 1$, $V_n^*(\tilde{c}_0, \cdot)$ *belong to the space* $\mathbb{B}_w([0, T] \times \mathbb{R} \times E)$ *and, moreover,*

$$\begin{aligned} &V_n^*(\tilde{c}_0, t, \lambda, x) \\ &\quad = \inf_{a \in A(x)} \biggl[\tilde{c}_0(t, \lambda, x, a) \\ &\qquad\qquad + \int_{[0,T]\times\mathbb{R}\times E} V_{n-1}^*(\tilde{c}_0, s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a) \biggr], \qquad n \ge 1, \end{aligned}$$

with $V_0^*(\tilde{c}_0, \cdot) := 0$ *and* $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$.

(ii) $\lim_{n\to\infty} V_n^*(\tilde{c}_0, t, \lambda, x) = V^*(\tilde{c}_0, t, \lambda, x)$ *for all* $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$.

(iii) $V^*(\tilde{c}_0, \cdot)$ *is the unique solution to the Bellman equation in the space* $\mathbb{B}_w([0, T] \times \mathbb{R} \times E)$, *i.e.*

$$v(t, \lambda, x) = \inf_{a \in A(x)} \biggl[\tilde{c}_0(t, \lambda, x, a) + \int_{[0,T]\times\mathbb{R}\times E} v(s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a) \biggr]$$

for $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$.

(iv) *There exists a policy* $\phi^* \in \Pi^{\mathrm{DS}}$ *such that*

$$\begin{aligned} &\tilde{c}_0(t, \lambda, x, \phi^*(t, \lambda, x)) + \int_{[0,T]\times\mathbb{R}\times E} V^*(\tilde{c}_0, s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, \phi^*(t, \lambda, x)) \\ &\quad = \inf_{a \in A(x)} \biggl[\tilde{c}_0(t, \lambda, x, a) + \int_{[0,T]\times\mathbb{R}\times E} V^*(\tilde{c}_0, s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a) \biggr] \end{aligned}$$

for each $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$, *and such a policy* ϕ^* *is optimal for problem (3.8).*

Proof. Assumption 3.1(iii) implies that the discrete-time control model with the state space $[0, T] \times \mathbb{R} \times E$, the transition kernel $\mathbb{Q}$, and the one-stage cost $\tilde{c}_0$ is transient. Thus, the theory for transient MDPs (see, for example, [15, Section 9.6]) can be applied in our context. Note that Assumptions 3.1(i), 3.1(iii), and 4.1 verify [15, Assumption 9.6.8]. Hence, our main results here follow from [15, Lemma 9.6.9 and Theorem 9.6.10]. In fact, the policy $\phi^* \in \Pi^{\mathrm{DS}}$ in (iv) satisfies $V^{\phi^*}(\tilde{c}_0, t, \lambda, x) = V^*(\tilde{c}_0, t, \lambda, x)$ for all $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$, which, of course, implies that $V^{\phi^*}(\tilde{c}_0, \nu) = V^*(\tilde{c}_0, \nu)$ and, thus, ϕ^* is optimal for problem (3.8). □

Algorithm 4.1. In Theorem 4.1 we proposed an algorithm for computing optimal policies:

(i) approximate $V^*(\tilde{c}_0, \cdot)$ by iterating $V_n^*(\tilde{c}_0, \cdot)$ over n until n is large enough;

(ii) compute the minimum point of the right-hand side of the Bellman equation over $A(x)$ for all (t, λ, x), denoted by $\phi^*(t, \lambda, x)$, which is an optimal policy.

Remark 4.2. In the above algorithm, the augmented state variables and the action variable are generally continuous. To implement the algorithm in practice, one should first discretize the continuous variables, i.e. one should partition the continuous spaces into finite parts with suitable scales; see, for example, [1, Section 7.5.3]. Nonetheless, the computation will be still very complex.

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

5. Solution to the constrained problem

5.1. Existence of constrained-optimal policies

From our preliminary analysis in Section 3 we see that, under Assumption 3.1, the constrained problem (3.9) can be written as the following convex program in $\mathcal{D} := \{\eta_\nu^\pi : \pi \in \Pi\}$:

$$\begin{aligned} &\min_{\eta\in\mathcal{D}} && \int_{\mathbb{K}} \tilde{c}_0(s,\gamma,y,a)\eta(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y,\,\mathrm{d}a) \\ &\text{subject to} && \int_{\mathbb{K}} \tilde{c}_i(s,\gamma,y,a)\eta(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y,\,\mathrm{d}a) \le d_i, \qquad i = 1,\dots,q. \end{aligned} \tag{5.1}$$

This convex program has at least one feasible solution due to the assumption that $\mathbb{U} \neq \varnothing$. To ensure that this convex program is solvable, however, we need to establish the compactness of the measure space $\mathcal{D}_0 := \{\eta \in \mathcal{D} \colon \int_{\mathbb{K}}\tilde{c}_i\,\mathrm{d}\eta \le d_i,\ i = 1,\dots,q\}$ and the continuity of $\int_{\mathbb{K}}\tilde{c}_0\,\mathrm{d}\eta$ in $\eta \in \mathcal{D}_0$ in some sense. Because the cost functions $\tilde{c}_i$ have finite w-norms, we first introduce some measure spaces with the so-called finite w-norm and associated weak topologies.

Definition 5.1. (i) A signed measure η on $\mathbb{K}$ or on $[0,T]\times\mathbb{R}\times E$ is said to have a finite w-norm if $\int_{\mathbb{K}} w(t,\lambda,x)|\eta|(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a) < \infty$ or $\int_{[0,T]\times\mathbb{R}\times E} w(t,\lambda,x)|\eta|(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x) < \infty$, respectively, where $|\eta|$ denotes the total variation of η. The spaces of signed measures on $\mathbb{K}$ and on $[0,T]\times\mathbb{R}\times E$ with finite w-norms are denoted by $\mathbb{M}_w(\mathbb{K})$ and $\mathbb{M}_w([0,T]\times\mathbb{R}\times E)$, respectively. Similarly, the spaces of nonnegative measures on $\mathbb{K}$ and on $[0,T]\times\mathbb{R}\times E$ with finite w-norms are denoted by $\mathbb{M}_w^+(\mathbb{K})$ and $\mathbb{M}_w^+([0,T]\times\mathbb{R}\times E)$, respectively.

(ii) The w-weak topology on $\mathbb{M}_w(\mathbb{K})$ is the weakest topology on $\mathbb{M}_w(\mathbb{K})$ such that $\int_{\mathbb{K}} v\,\mathrm{d}\eta$ is continuous in $\eta \in \mathbb{M}_w(\mathbb{K})$ for each continuous $v \in \mathbb{B}_w(\mathbb{K})$. This topology is denoted by $\sigma(\mathbb{M}_w(\mathbb{K}))$, and the corresponding weak convergence is represented by '$\xrightarrow{\mathrm{w}}$'. Similarly, $\sigma(\mathbb{M}_w^+(\mathbb{K}))$ is the weakest topology on $\mathbb{M}_w^+(\mathbb{K})$ such that $\int_{\mathbb{K}} v\,\mathrm{d}\eta$ is continuous in $\eta \in \mathbb{M}_w^+(\mathbb{K})$ for each continuous $v \in \mathbb{B}_w(\mathbb{K})$.

Note that, when $w \equiv 1$, $\mathbb{B}_w(\mathbb{K})$ is the space of bounded measurable functions, $\mathbb{M}_w(\mathbb{K})$ is the space of finite signed measures, and $\sigma(\mathbb{M}_w(\mathbb{K}))$ is the usual weak topology. In this case, we omit w from the notation for simplicity. To establish the relationship between the w-weak convergence in $(\mathbb{M}_w^+(\mathbb{K}), \sigma(\mathbb{M}_w^+(\mathbb{K})))$ and the usual weak convergence in $(\mathbb{M}^+(\mathbb{K}), \sigma(\mathbb{M}^+(\mathbb{K})))$, we define two operators, H_w on $\mathbb{M}_w^+(\mathbb{K})$ and H_w^{-1} on $\mathbb{M}^+(\mathbb{K})$, as follows: for $\eta \in \mathbb{M}_w^+(\mathbb{K})$, $\mu \in \mathbb{M}^+(\mathbb{K})$, and $\Gamma \in \mathcal{B}(\mathbb{K})$,

$$H_w\eta(\Gamma) := \int_\Gamma w(t,\lambda,x)\eta(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a) \quad \text{and} \quad H_w^{-1}\mu(\Gamma) := \int_\Gamma \frac{\mu(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a)}{w(t,\lambda,x)}.$$

Lemma 5.1. *Let $w \ge 1$ be a continuous function on $[0,T]\times\mathbb{R}\times E$. Then the following hold:*

(i) $H_w\eta \in \mathbb{M}^+(\mathbb{K})$ *for all* $\eta \in \mathbb{M}_w^+(\mathbb{K})$, $H_w^{-1}\mu \in \mathbb{M}_w^+(\mathbb{K})$ *for all* $\mu \in \mathbb{M}^+(\mathbb{K})$;

(ii) $H_w^{-1}(H_w\eta) = \eta$ *for all* $\eta \in \mathbb{M}_w^+(\mathbb{K})$, $H_w(H_w^{-1}\mu) = \mu$ *for all* $\mu \in \mathbb{M}^+(\mathbb{K})$;

(iii) $\eta_k \xrightarrow{\mathrm{w}} \eta$ *if and only if* $H_w\eta_k \xrightarrow{1} H_w\eta$, $\mu_k \xrightarrow{1} \mu$ *if and only if* $H_w^{-1}\mu_k \xrightarrow{\mathrm{w}} H_w^{-1}\mu$.

Proof. The results follow from [11, Lemma 8] or [18, Lemma 3.4]. □

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

To establish the compactness of the space $\mathcal{D}$ in $(\mathbb{M}_w^+(\mathbb{K}), \sigma(\mathbb{M}_w^+(\mathbb{K})))$ and the continuity of $\int_{\mathbb{K}} \tilde{c}_i \, \mathrm{d}\eta$ in $\eta \in \mathcal{D}$, we propose another set of compactness-continuity conditions.

Assumption 5.1. (i) *The function w from Assumption 3.1 is continuous on $[0, T] \times \mathbb{R} \times E$.*

(ii) *There exist a continuous function $\bar{w}$ on $[0, T] \times \mathbb{R} \times E$ and an increasing sequence of compact sets $\mathbb{K}_m \uparrow \mathbb{K}$ as $m \uparrow \infty$ such that*

$$\lim_{m\to\infty} \inf_{(t,\lambda,x,a)\in\mathbb{K}\setminus\mathbb{K}_m} \frac{\bar{w}(t,\lambda,x)}{w(t,\lambda,x)} = \infty \quad \textit{and} \quad \sup_{\eta\in\mathcal{D}} \int_{\mathbb{K}} \bar{w}(t,\lambda,x)\eta(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a) < \infty.$$

(iii) $\int_0^{T-t} \int_E v(t+s, \lambda + \int_t^{t+s} c(z,x,a)\,\mathrm{d}z, y) Q(\mathrm{d}s, \mathrm{d}y \mid x, a)$ *is continuous in $(t, \lambda, x, a) \in \mathbb{K}$ for each bounded continuous function v on $[0, T] \times \mathbb{R} \times E$.*

(iv) *$Q(T - t, E \mid x, a)$ is upper semicontinuous in $(t, x, a) \in [0, T] \times K$.*

(v) *$c(t, x, a)$ is lower semicontinuous in $(t, x, a) \in [0, T] \times K$.*

Remark 5.1. Assumptions 5.1(iv) and 5.1(v) along with the properties of u_i imply that $\tilde{c}_i(t, \lambda, x, a)$ is lower semicontinuous in $(t, \lambda, x, a) \in \mathbb{K}$ for each $i = 0, 1, \ldots, q$.

We can now state the existence of optimal policies for the constrained problem (3.9).

Theorem 5.1. *Suppose that Assumptions 3.1(ii), 3.1(iii), and 5.1(i)–5.1(iii) hold. Then the following assertions hold:*

(i) *the space of occupation measure $\mathcal{D}$ is compact in $(\mathbb{M}_w^+(\mathbb{K}), \sigma(\mathbb{M}_w^+(\mathbb{K})))$;*

(ii) *if, furthermore, Assumption 3.1(i), and Assumptions 5.1(iv) and 5.1(v) are satisfied, then there exists an optimal solution to problem (5.1) and, thus, there is a randomized jump-stationary optimal policy for the constrained problem (3.9).*

Proof. (i) It is clear that $\mathcal{D}$ is a subset of $\mathbb{M}_w^+(\mathbb{K})$ under Assumptions 3.1(ii) and 3.1(iii). We now prove that $\mathcal{D}$ is relatively compact in $(\mathbb{M}_w^+(\mathbb{K}), \sigma(\mathbb{M}_w^+(\mathbb{K})))$. Since w is assumed to be continuous, by Lemma 5.1, we need only show that $\tilde{\mathcal{D}} := \{\mu := H_w\eta \colon \eta \in \mathcal{D}\}$ is relatively compact in $(\mathbb{M}^+(\mathbb{K}), \sigma(\mathbb{M}^+(\mathbb{K})))$. Indeed, on the one hand, for the moment function $\bar{w}/w$ from Assumption 5.1(ii), it holds that

$$\sup_{\mu\in\tilde{\mathcal{D}}} \int_{\mathbb{K}} \frac{\bar{w}(t,\lambda,x)}{w(t,\lambda,x)} \mu(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a) = \sup_{\eta\in\mathcal{D}} \int_{\mathbb{K}} \bar{w}(t,\lambda,x)\eta(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a) < \infty,$$

which, by [15, Theorem 12.2.15], implies that $\tilde{\mathcal{D}}$ is tight. On the other hand, it follows from Assumptions 3.1(ii) and 3.1(iii) that $\sup_{\mu\in\tilde{\mathcal{D}}} \mu(\mathbb{K}) = \sup_{\eta\in\mathcal{D}} \int_{\mathbb{K}} w \, \mathrm{d}\eta < \infty$. Thus, by [11, Lemma 7], $\tilde{\mathcal{D}}$ is relatively compact in $(\mathbb{M}^+(\mathbb{K}), \sigma(\mathbb{M}^+(\mathbb{K})))$.

To complete the proof, we still need to show that $\mathcal{D}$ is closed in $(\mathbb{M}_w^+(\mathbb{K}), \sigma(\mathbb{M}_w^+(\mathbb{K})))$. To do so, let $\{\eta_n, n \geq 0\} \subset \mathcal{D}$ be a sequence such that $\eta_n \xrightarrow{\mathrm{W}} \eta$. By [11, Lemma 6], $\mathbb{M}_w^+(\mathbb{K})$ is closed in $(\mathbb{M}_w(\mathbb{K}), \sigma(\mathbb{M}_w(\mathbb{K})))$ and, therefore, $\eta \in \mathbb{M}_w^+(\mathbb{K})$. To further show that $\eta \in \mathcal{D}$, by Theorem 3.1, we need only prove that η satisfies (3.4). However, this follows from Assumption 5.1(iii) and the fact that $\{\eta_n, n \geq 0\} \subset \mathcal{D}$. Thus, $\mathcal{D}$ is closed in $(\mathbb{M}_w^+(\mathbb{K}), \sigma(\mathbb{M}_w^+(\mathbb{K})))$, which, together with the relative compactness of $\mathcal{D}$, guarantees the compactness of $\mathcal{D}$.

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

(ii) As shown above, problem (5.1) can be written as $\min_{\eta\in\mathcal{D}_0}\int_{\mathbb{K}}\tilde{c}_0\,\mathrm{d}\eta$. When Assumptions 3.1(i), 5.1(iv), and 5.1(v) are further imposed, we can show that for each $i=1,\dots,q$, $\int_{\mathbb{K}}\tilde{c}_i\,\mathrm{d}\eta$ is lower semicontinuous in $\eta\in\mathcal{D}$ with the w-weak topology. Thus, the space $\mathcal{D}_0$ is closed in $\mathcal{D}$ and, therefore, by part (i), $\mathcal{D}_0$ is compact. This fact, together with the lower semicontinuity of $\int_{\mathbb{K}}\tilde{c}_0\,\mathrm{d}\eta$ in $\eta\in\mathcal{D}_0$, guarantees the existence of an optimal solution η^* to problem (5.1). By Theorem 3.1, the policy $\phi^*\in\Pi^{\mathrm{RS}}$ determined by $\eta^*(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a)=\phi^*(\mathrm{d}a\mid t,\lambda,x)\hat{\eta}^*(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x)$ is optimal for the constrained problem (3.9). □

5.2. Linear programming formulation and strong duality

In this subsection we present a linear programming formulation of the constrained problem (3.9), develop the dual program, and also establish the absence of the duality gap.

First, by Theorem 3.1, problem (5.1) is equivalent to the following linear program (LP):

$$\begin{aligned}
\min_{\eta}\quad & \int_{\mathbb{K}}\tilde{c}_0(s,\gamma,y,a)\eta(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y,\,\mathrm{d}a)\\
\text{subject to}\quad & \hat{\eta}(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y)=\nu(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y)\\
&\qquad\qquad +\int_{\mathbb{K}}\mathbb{Q}(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y\mid t,\lambda,x,a)\eta(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a),\\
& \int_{\mathbb{K}}\tilde{c}_i(s,\gamma,y,a)\eta(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y,\,\mathrm{d}a)\le d_i,\qquad i=1,\dots,q,\\
& \eta\in\mathbb{M}_w^+(\mathbb{K}).
\end{aligned}\tag{5.2}$$

Under Assumptions 3.1 and 5.1, we know from Theorem 5.1 that the LP (5.2) is solvable. If η^* is an optimal solution to problem (5.2), by Theorem 5.1, the policy $\phi^*\in\Pi^{\mathrm{RS}}$ determined by $\eta^*(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a)=\phi^*(\mathrm{d}a\mid t,\lambda,x)\hat{\eta}^*(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x)$ is optimal for the constrained problem (3.9).

Remark 5.2. In general, the LP (5.2) is an infinite-dimensional linear program with infinitely many (uncountable) variables. In practical computation, such a linear program cannot be directly implemented in a computer. The alternative way is to approximate the infinite-dimensional linear program by finite-dimensional linear programs with finitely many variables. There are two approaches for making finite approximations in the literature. One is the aggregation-relaxation-inner approximation method; see [13, Section 4.1] or [15, Section 12.5] for details. Another is to discretize the augmented state variables; see [13, Section 4.2] for details.

We next derive the dual program of the primal program (5.2). Unlike the primal program in occupation measures, the dual program is a linear program in value functions. To proceed, we introduce some notation. First, $(\mathbb{M}_w(\mathbb{K}),\mathbb{B}_w(\mathbb{K}))$ become a dual pair if we define the bilinear form on the pair by

$$\langle\eta,h\rangle:=\int_{\mathbb{K}}h(t,\lambda,x,a)\eta(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a)\quad\text{for all }\eta\in\mathbb{M}_w(\mathbb{K}),\ h\in\mathbb{B}_w(\mathbb{K}).$$

Similarly, we define the bilinear form on the pair $(\mathbb{M}_w([0,T]\times\mathbb{R}\times E),\mathbb{B}_w([0,T]\times\mathbb{R}\times E))$. Moreover, we define the bilinear forms on $(\mathbb{R}^q,\mathbb{R}^q)$ and $(\mathbb{M}_w(\mathbb{K})\times\mathbb{R}^q,\mathbb{B}_w(\mathbb{K})\times\mathbb{R}^q)$ by

$$\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle:=\sum_{i=1}^q\alpha_i\beta_i,\quad\text{and}\quad\langle(\eta,\boldsymbol{\alpha}),(h,\boldsymbol{\beta})\rangle:=\int_{\mathbb{K}}h(t,\lambda,x,a)\eta(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x,\,\mathrm{d}a)+\sum_{i=1}^q\alpha_i\beta_i,$$

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

respectively, with $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q) \in \mathbb{R}^q$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q) \in \mathbb{R}^q$, $\eta \in \mathbb{M}_w(\mathbb{K})$, and $h \in \mathbb{B}_w(\mathbb{K})$. Now, let L_0 be the linear map on $\mathbb{M}_w(\mathbb{K})$ defined by

$$(L_0\eta)(B, C, D) = \hat{\eta}(B, C, D) - \int_{\mathbb{K}} \mathbb{Q}(B, C, D \mid t, \lambda, x, a)\eta(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a).$$

Under Assumption 3.1(iii), it is clear that L_0 maps $\mathbb{M}_w(\mathbb{K})$ to $\mathbb{M}_w([0, T] \times \mathbb{R} \times E)$. Further, we introduce another linear map $L\colon \mathbb{M}_w(\mathbb{K}) \times \mathbb{R}^q \to \mathbb{M}_w([0, T] \times \mathbb{R} \times E) \times \mathbb{R}^q$ as

$$L(\eta, \boldsymbol{\alpha}) := (L_0\eta, \langle \eta, \tilde{c}_1\rangle + \alpha_1, \dots, \langle \eta, \tilde{c}_q\rangle + \alpha_q) \quad \text{for all } (\eta, \boldsymbol{\alpha}) \in \mathbb{M}_w(\mathbb{K}) \times \mathbb{R}^q.$$

Using the notation above, we can restate the LP (5.2) as

$$\begin{aligned} \min_{\eta, \boldsymbol{\alpha}} \quad & \langle (\eta, \boldsymbol{\alpha}), (\tilde{c}_0, \mathbf{0})\rangle \\ \text{subject to} \quad & L(\eta, \boldsymbol{\alpha}) = (\nu, \boldsymbol{d}), \\ & (\eta, \boldsymbol{\alpha}) \in \mathbb{M}_w^+(\mathbb{K}) \times \mathbb{R}_+^q. \end{aligned} \tag{5.3}$$

Here, we denote $\boldsymbol{d} = (d_1, \dots, d_q)$.

Theorem 5.2. *The adjoint* $L^*\colon \mathbb{B}_w([0, T] \times \mathbb{R} \times E) \times \mathbb{R}^q \to \mathbb{B}_w(\mathbb{K}) \times \mathbb{R}^q$ *of* L *is*

$$L^*(v, \boldsymbol{\beta}) = \bigg(L_0^* v + \sum_{i=1}^q \beta_i \tilde{c}_i, \boldsymbol{\beta}\bigg) \quad \textit{for all } (v, \boldsymbol{\beta}) \in \mathbb{B}_w([0, T] \times \mathbb{R} \times E) \times \mathbb{R}^q,$$

where $L_0^*\colon \mathbb{B}_w([0, T] \times \mathbb{R} \times E) \to \mathbb{B}_w(\mathbb{K})$ *is the adjoint of* L_0 *defined by*

$$\begin{aligned} L_0^* v(t, \lambda, x, a) &:= v(t, \lambda, x) \\ &\quad - \int_{[0,T]\times\mathbb{R}\times E} v(s, \gamma, y)\mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a), (t, \lambda, x, a) \in \mathbb{K}. \end{aligned} \tag{5.4}$$

Proof. We first derive the adjoint L_0^* of L_0. From the definition, we deduce that

$$\begin{aligned} \langle L_0\eta, v\rangle &= \int_{[0,T]\times\mathbb{R}\times E} v(s, \gamma, y) \bigg[\hat{\eta}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y) \\ &\qquad\qquad - \int_{\mathbb{K}} \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a)\eta(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a)\bigg] \\ &= \int_{\mathbb{K}} \bigg[v(t, \lambda, x) \\ &\qquad - \int_{[0,T]\times\mathbb{R}\times E} v(s, \gamma, y)\mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a)\bigg]\eta(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a) \\ &= \langle \eta, L_0^* v\rangle, \end{aligned}$$

which implies that L_0^* has the form of (5.4). Now, the adjoint L^* of L is derived from a direct calculation as

$$\begin{aligned} \langle L(\eta, \boldsymbol{\alpha}), (v, \boldsymbol{\beta})\rangle &= \langle L_0\eta, v\rangle + \sum_{i=1}^q \beta_i(\langle \eta, \tilde{c}_i\rangle + \alpha_i) \\ &= \langle \eta, L_0^* v\rangle + \bigg\langle \eta, \sum_{i=1}^q \beta_i \tilde{c}_i \bigg\rangle + \sum_{i=1}^q \alpha_i \beta_i \end{aligned}$$

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

$$= \left\langle \eta, L_0^* v + \sum_{i=1}^{q} \beta_i \tilde{c}_i \right\rangle + \sum_{i=1}^{q} \alpha_i \beta_i$$
$$= \left\langle (\eta, \boldsymbol{\alpha}), \left(L_0^* v + \sum_{i=1}^{q} \beta_i \tilde{c}_i, \boldsymbol{\beta} \right) \right\rangle$$
$$= \langle (\eta, \boldsymbol{\alpha}), L^*(v, \boldsymbol{\beta}) \rangle. \qquad \square$$

According to [15, Chapter 6], the dual program of the primal LP (5.3) is of the form

$$\begin{aligned} \max_{v,\boldsymbol{\beta}} \quad & \langle (\nu, \boldsymbol{d}), (v, \boldsymbol{\beta}) \rangle \\ \text{subject to} \quad & (\tilde{c}_0, \mathbf{0}) - L^*(v, \boldsymbol{\beta}) \in \mathbb{B}_w^+(\mathbb{K}) \times \mathbb{R}_+^q, \\ & (v, \boldsymbol{\beta}) \in \mathbb{B}_w([0, T] \times \mathbb{R} \times E) \times \mathbb{R}^q. \end{aligned} \tag{5.5}$$

The more explicit form is

$$\begin{aligned} \max_{v,\boldsymbol{\beta}} \quad & \int_{[0,T]\times\mathbb{R}\times E} v(t, \lambda, x) \nu(\mathrm{d}t, \mathrm{d}\lambda, \mathrm{d}x) - \sum_{i=1}^{q} \beta_i d_i, \\ \text{subject to} \quad & \tilde{c}_0(t, \lambda, x, a) - v(t, \lambda, x) \\ & \quad + \int_{[0,T]\times\mathbb{R}\times E} v(s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a) \\ & \quad + \sum_{i=1}^{q} \beta_i \tilde{c}_i(t, \lambda, x, a) \\ & \quad \geq 0 \quad \text{for all } (t, \lambda, x, a) \in \mathbb{K}, \\ & \quad v \in \mathbb{B}_w([0, T] \times \mathbb{R} \times E), \beta_i \geq 0, \ i = 1, \dots, q, \end{aligned} \tag{5.6}$$

where we have replaced β_i with $-\beta_i$.

Remark 5.3. According to the generalized Farkas theorem (see [15, Theorem 12.2.11]), the LP (5.2) has a feasible solution (and, thus, $\mathbb{U} \neq \varnothing$) if and only if, for any $(v, \boldsymbol{\beta}) \in \mathbb{B}_w([0, T] \times \mathbb{R} \times E) \times \mathbb{R}^q$ such that $L^*(v, \boldsymbol{\beta}) \in \mathbb{B}_w^+(\mathbb{K}) \times \mathbb{R}_+^q$, we have $\langle (\nu, \boldsymbol{d}), (v, \boldsymbol{\beta}) \rangle \geq 0$.

Below, we denote the values of the primal LP (5.2) and the dual LP (5.6) by inf(LP) and sup(DP), respectively. To establish the strong duality, we also need the Slater-like condition.

Assumption 5.2. *There exists a policy* $\pi \in \Pi$ *such that* $V^\pi(\tilde{c}_i, \nu) < d_i$, $i = 1, 2, \dots, q$.

Remark 5.4. The Slater condition is widely used in constrained optimization problems as a sufficient condition to ensure no duality gap; see [4] and [20]. Assumption 5.2 is a Slater-like condition in our setup. To check it in practice, one possible option may be to compute $V^\pi(\tilde{c}_i, \nu)$ under some special policy π and then check to see if $V^\pi(\tilde{c}_i, \nu) < d_i$. For example, for a given policy $\phi \in \Pi^{\mathrm{DS}}$, under the condition of Theorem 4.1, $V^\phi(\tilde{c}_i, t, \lambda, x)$ can be computed as the unique solution in the space $\mathbb{B}_w([0, T] \times \mathbb{R} \times E)$ to the following equation:

$$v(t, \lambda, x) = \tilde{c}_0(t, \lambda, x, \phi(t, \lambda, x)) + \int_{[0,T]\times\mathbb{R}\times E} v(s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, \phi(t, \lambda, x))$$

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

for $(t, \lambda, x) \in [0, T] \times \mathbb{R} \times E$. Since the variables t, λ, and x are all continuous in our setup, it is difficult to solve this equation. An alternative way to solve the equation may be the numerical iterative algorithm; see Remark 4.2 for more details.

We now state the strong duality theorem.

Theorem 5.3. *Under Assumptions 3.1–5.2, the strong duality between the primal LP (5.2) and its dual LP (5.6) holds, i.e. both problems admit optimal solutions and* $\inf(\mathrm{LP}) = \sup(\mathrm{DP})$.

Proof. In Theorem 5.1, we proved that $\mathcal{D}$ is w-weakly closed in $\mathbb{M}_w^+(\mathbb{K})$, and for $i = 1, \dots, q$, $\int_{\mathbb{K}} \tilde{c}_i \, \mathrm{d}\eta$ is w-weakly lower semicontinuous in $\eta \in \mathcal{D}$. Since the form of the LP (5.2) is the same as the formulation of [20, Example 1], under Assumptions 3.1–5.2, it follows from [20, Theorem 17 and Example 1, pp. 7, 18, 23] that there exist constants $\beta_i^* \geq 0$ $(i = 1, \dots, q)$ such that

$$\begin{aligned}\inf(\mathrm{LP}) &= \sup_{\beta_i \geq 0,\, i=1,\dots,q} \inf_{\eta \in \mathcal{D}} \bigg[\int_{\mathbb{K}} \bigg(\tilde{c}_0(s, \gamma, y, a) + \sum_{i=1}^{q} \beta_i \tilde{c}_i(s, \gamma, y, a) \bigg) \eta(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y, \mathrm{d}a) \\ &\qquad\qquad - \sum_{i=1}^{q} \beta_i d_i \bigg] \\ &= \inf_{\eta \in \mathcal{D}} \bigg[\int_{\mathbb{K}} \bigg(\tilde{c}_0(s, \gamma, y, a) + \sum_{i=1}^{q} \beta_i^* \tilde{c}_i(s, \gamma, y, a) \bigg) \eta(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y, \mathrm{d}a) - \sum_{i=1}^{q} \beta_i^* d_i \bigg].\end{aligned} \tag{5.7}$$

Now, for arbitrarily fixed $\beta_1, \dots, \beta_q$, we can prove, by Theorem 4.1 (with $(\tilde{c}_0 + \sum_{i=1}^{q} \beta_i \tilde{c}_i)$ in lieu of $\tilde{c}_0$) and by contradiction, that

$$\begin{aligned}&\inf_{\eta \in \mathcal{D}} \bigg[\int_{\mathbb{K}} \bigg(\tilde{c}_0(s, \gamma, y, a) + \sum_{i=1}^{q} \beta_i \tilde{c}_i(s, \gamma, y, a) \bigg) \eta(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y, \mathrm{d}a) \bigg] \\ &\quad = \sup_{v \in \mathbb{B}_w([0,T] \times \mathbb{R} \times E)} \bigg[\int_{[0,T] \times \mathbb{R} \times E} v(s, \gamma, y) \nu(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y) \colon \tilde{c}_0(t, \lambda, x, a) \\ &\qquad\qquad + \sum_{i=1}^{q} \beta_i \tilde{c}_i(t, \lambda, x, a) \\ &\qquad\qquad + \int_{[0,T] \times \mathbb{R} \times E} v(s, \gamma, y) \mathbb{Q}(\mathrm{d}s, \mathrm{d}\gamma, \mathrm{d}y \mid t, \lambda, x, a) \\ &\qquad\qquad \geq v(t, \lambda, x) \quad \text{for all } (t, \lambda, x, a) \in \mathbb{K} \bigg].\end{aligned} \tag{5.8}$$

By (5.8), we see that $\sup(\mathrm{DP})$ coincides with the term on the right-hand side of (5.7). Therefore, it follows from (5.7) that $\inf(\mathrm{LP}) = \sup(\mathrm{DP})$. The solvability of the primal LP (5.2) is guaranteed by Theorem 5.1, whereas an optimal solution for the dual LP (5.6) is given by $(V^*(\tilde{c}_0 + \sum_{i=1}^{q} \beta_i^* \tilde{c}_i, \cdot), \beta_1^*, \dots, \beta_q^*)$. □

Remark 5.5. (i) To see the relationship between the primal LP (5.3) and the dual LP (5.5) more explicitly, we let the Lagrange multiplier for constraint $L_0 \eta = \nu$ be $v \in \mathbb{B}_w([0, T] \times \mathbb{R} \times E)$,

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

and the Lagrange multiplier for constraint $(\langle\eta,\tilde{c}_1\rangle+\alpha_1,\dots,\langle\eta,\tilde{c}_q\rangle+\alpha_q)=(d_1,\dots,d_q)$ be $\boldsymbol{\beta}\in\mathbb{R}^q$. Then, the Lagrangian for the primal LP (5.3) is

$$\Psi(\eta,\boldsymbol{\alpha},v,\boldsymbol{\beta})=\langle\eta,\tilde{c}_0\rangle+\langle L_0\eta-\nu,v\rangle+\langle(\langle\eta,\tilde{c}_1\rangle+\alpha_1-d_1,\dots,\langle\eta,\tilde{c}_q\rangle+\alpha_q-d_q),\boldsymbol{\beta}\rangle.$$

The LP (5.3) is equivalent to $\inf_{\eta\geq 0,\,\boldsymbol{\alpha}\geq 0}\sup_{v,\boldsymbol{\beta}}\Psi(\eta,\boldsymbol{\alpha},v,\boldsymbol{\beta})$, the dual problem of which is defined to be $\sup_{v,\boldsymbol{\beta}}\inf_{\eta\geq 0,\,\boldsymbol{\alpha}\geq 0}\Psi(\eta,\boldsymbol{\alpha},v,\boldsymbol{\beta})$. Rearranging the Lagrangian yields

$$\Psi(\eta,\boldsymbol{\alpha},v,\boldsymbol{\beta})=-\langle\nu,v\rangle-\langle\boldsymbol{d},\boldsymbol{\beta}\rangle+\left\langle\eta,\tilde{c}_0+L_0^*v+\sum_{i=1}^{q}\beta_i\tilde{c}_i\right\rangle+\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle,$$

which leads to the dual LP (5.5). Hence, the variables v and $\boldsymbol{\beta}$ in the dual LP (5.5) are just the Lagrange multipliers for constraints in the primal LP (5.3).

(ii) The dual LP aids in calculating the value function. In general, we cannot derive a constrained-optimal policy from the dual LP, unlike the case of the primal LP in occupation measures. Formally, under the assumptions in Theorem 5.3, the dual LP (5.5) has an optimal solution $(V^*(\tilde{c}_0+\sum_{i=1}^{q}\beta_i^*\tilde{c}_i,\cdot),\ \beta_1^*,\dots,\beta_q^*)$, from which we can deduce a policy $\phi^*\in\Pi^{\mathrm{DS}}$ such that

$$\begin{aligned}&V^{\phi^*}\left(\tilde{c}_0+\sum_{i=1}^{q}\beta_i^*\tilde{c}_i,t,\lambda,x\right)\\&\quad=V^*\left(\tilde{c}_0+\sum_{i=1}^{q}\beta_i^*\tilde{c}_i,t,\lambda,x\right)\quad\text{for all }(t,\lambda,x)\in[0,T]\times\mathbb{R}\times E,\end{aligned}$$

but we do not know if it is constrained-optimal. However, in the special case of $\mathbb{U}=\Pi$, the dual LP (5.6) is reduced to the linear programming formulation for the unconstrained problem (3.9), i.e.

$$\begin{aligned}\max_{v,\boldsymbol{\beta}}\quad&\int_{[0,T]\times\mathbb{R}\times E}v(t,\lambda,x)\nu(\mathrm{d}t,\,\mathrm{d}\lambda,\,\mathrm{d}x)\\\text{subject to}\quad&\tilde{c}_0(t,\lambda,x,a)-v(t,\lambda,x)\\&\quad+\int_{[0,T]\times\mathbb{R}\times E}v(s,\gamma,y)\mathbb{Q}(\mathrm{d}s,\,\mathrm{d}\gamma,\,\mathrm{d}y\mid t,\lambda,x,a)\\&\quad\geq 0\quad\text{for all }(t,\lambda,x,a)\in\mathbb{K},\ v\in\mathbb{B}_w([0,T]\times\mathbb{R}\times E),\end{aligned}$$

which produces the optimal solution $V^*(\tilde{c}_0,\cdot)$, based on which an optimal policy can be derived in the same way as in Theorem 4.1(iv).

6. Some special cases

In this section we provide two special cases of our results. One is on the discrete-time case, another is about the chance-constrained problem.

6.1. Special case 1: discrete-time case

As is known, when the sojourn times between jumps are all equal to 1, an SMDP will be reduced to a DTMDP. Here, we consider this situation and, for simplicity, assume that the initial

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

time $T_0 \equiv 0$ and the costs occur with invariable rates $c(n, x, a)$ over $[n, n+1)$. In this way, $\Lambda_{T+1} = \Lambda_T$ and the expected utilities are reduced to the form

$$\mathbb{J}_i^\pi(\nu) = \mathbb{E}_\nu^\pi[u_i(\Lambda_T)] = \mathbb{E}_\nu^\pi\biggl[u_i\biggl(\Lambda_0 + \sum_{n=0}^{T-1} c(n, X_n, A_n)\biggr)\biggr] \quad \text{for all } \pi \in \Pi,\ i = 0, \dots, q,$$

where the ending time T is now assumed to be an integer. The occupation measure of a policy π in (3.3) is then reduced to the form

$$\eta_\nu^\pi(n, C, D, \Gamma) = \mathbb{P}_\nu^\pi(\Lambda_n \in C, X_n \in D, A_n \in \Gamma), \qquad n = 0, 1, \dots, T.$$

From this, we see that our occupation measures differ from those of [13]. Moreover, the distribution of Λ_T is of the form $\mathbb{P}_\nu^\pi(\Lambda_T \in D) = \eta_\nu^\pi(T, D, E, A)$ for $D \in \mathcal{B}(\mathbb{R})$ and, therefore, the expected utilities can be expressed as

$$\mathbb{J}_i^\pi(\nu) = \int_{\mathbb{R}} u_i(\gamma)\eta_\nu^\pi(T, \mathrm{d}\gamma, E, A) \quad \text{for all } \pi \in \Pi,\ i = 0, \dots, q.$$

Hence, the constrained problem (3.2) can be written as

$$\begin{aligned} \min_{\eta \in \mathcal{D}} \quad & \int_{\mathbb{R}} u_0(\gamma)\eta(T, \mathrm{d}\gamma, E, A) \\ \text{subject to} \quad & \int_{\mathbb{R}} u_i(\gamma)\eta(T, \mathrm{d}\gamma, E, A) \le d_i, \qquad i = 1, \dots, q, \end{aligned}$$

which, by Theorem 3.1, is equivalent to the LP

$$\begin{aligned} \min_{\eta} \quad & \int_{\mathbb{R}} u_0(\gamma)\eta(T, \mathrm{d}\gamma, E, A) \\ \text{subject to} \quad & \hat{\eta}(0, C, D) = \nu(\{0\}, C, D), \\ & \hat{\eta}(m, C, D) = \int_{\mathbb{R}\times E\times A} \mathbb{Q}(m, C, D \mid m-1, \lambda, x, a)\eta(m-1, \mathrm{d}\lambda, \mathrm{d}x, \mathrm{d}a), \\ & \int_{\mathbb{R}} u_i(\gamma)\eta(T, \mathrm{d}\gamma, E, A) \le d_i, \qquad m = 1, \dots, T,\ i = 1, 2, \dots, q, \\ & \eta(n, C, D, \Gamma) \ge 0, \qquad n \in \{0, 1, \dots, T\},\ C \in \mathcal{B}(\mathbb{R}),\ D \in \mathcal{B}(E),\ \Gamma \in \mathcal{B}(A). \end{aligned}$$

6.2. Special case 2: chance-constrained problem

Chance-constrained problems are an interesting class of problems that impose constraints on the probabilities of some events. For example, someone may concern himself/herself with the probabilities that the total cost exceeds some particular levels, and would like to keep these probabilities lower than some tolerable bounds. The particular levels and tolerable bounds are selected individually. In our context, if we take $\Lambda_0 \equiv 0$, $u_0(\lambda) = \lambda$, and $u_i(\lambda) = \mathbf{1}_{[b_i, +\infty)}(\lambda)$ for some constants b_i and $i = 1, \dots, q$, our problem (3.2) is reduced to a chance-constrained problem, i.e.

$$\begin{aligned} \min_{\pi \in \Pi} \quad & \mathbb{E}_\nu^\pi\biggl[\int_{T_0}^{T} c(s, \xi_s, \vartheta_s)\,\mathrm{d}s\biggr] \\ \text{subject to} \quad & \mathbb{P}_\nu^\pi\biggl(\int_{T_0}^{T} c(s, \xi_s, \vartheta_s)\,\mathrm{d}s \ge b_i\biggr) \le d_i, \qquad i = 1, \dots, q. \end{aligned}$$

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press

The equivalent linear programming formulation is as follows:

$$
\begin{aligned}
\min_{\eta} \quad & \int_{\mathbb{K}} \tilde{c}_0(s, \gamma, y, a)\eta(\mathrm{d}s,\, \mathrm{d}\gamma,\, \mathrm{d}y,\, \mathrm{d}a) \\
\text{subject to} \quad & \hat{\eta}(\mathrm{d}s,\, \mathrm{d}\gamma,\, \mathrm{d}y) = \nu(\mathrm{d}s,\, \mathrm{d}\gamma,\, \mathrm{d}y) \\
& \qquad\qquad + \int_{\mathbb{K}} \mathbb{Q}(\mathrm{d}s,\, \mathrm{d}\gamma,\, \mathrm{d}y \mid t, \lambda, x, a)\eta(\mathrm{d}t,\, \mathrm{d}\lambda,\, \mathrm{d}x,\, \mathrm{d}a), \\
& \int_{\mathbb{K}} \tilde{c}_i(s, \gamma, y, a)\eta(\mathrm{d}s,\, \mathrm{d}\gamma,\, \mathrm{d}y,\, \mathrm{d}a) \le d_i, \qquad i = 1, \dots, q,\ \eta \in \mathbb{M}_w^+(\mathbb{K}),
\end{aligned}
$$

where $\tilde{c}_i$ are now defined by

$$
\begin{aligned}
\tilde{c}_0(s, \gamma, y, a) &:= (1 - Q(T - s, E \mid y, a))\bigg(\gamma + \int_s^T c(z, y, a)\,\mathrm{d}z\bigg), \\
\tilde{c}_i(s, \gamma, y, a) &:= (1 - Q(T - s, E \mid y, a))\,\mathbf{1}_{\{\gamma + \int_s^T c(z,y,a)\,\mathrm{d}z \ge b_i\}}, \qquad i = 1, 2, \dots, q.
\end{aligned}
$$

Here, our results have generalized those of [13, Section 6.2] from a Markov model to a semi-Markov model.

7. Conclusion

Risk-sensitive control is an important issue in various economical and financial activities. In a general framework of finite-horizon SMDPs, we have studied the unconstrained and constrained risk-sensitive problems. The value iteration and linear programming procedures have been proposed to calculate unconstrained and constrained optimal policies, respectively. Since the augmented states are generally continuous, finite approximation techniques are needed in the practical computation. On the basis of the distribution of the finite-horizon cost in terms of occupation measures (Theorem 3.2), our analytic method can also be applied to deal with other risk-aware problems such as stochastic dominance-constrained optimization [12] in the context of finite-horizon SMDPs.

Acknowledgements

The corresponding author for this paper is Xianping Guo. This research was supported in part by the National Natural Science Foundation of China and the University of Macau (grant numbers MYRG2016-00016-FBA and FDCT/027/2016/A1). We are also grateful to the anonymous referees for their careful reading and many constructive suggestions that have improved this paper.

References

[1] Bäuerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer, Heidelberg.

[2] Bäuerle, N. and Rieder, U. (2014). More risk-sensitive Markov decision processes. *Math. Operat. Res.* **39,** 105–120.

[3] Beutler, F. J. and Ross, K. W. (1986). Time-average optimal constrained semi-Markov decision processes. *Adv. Appl. Prob.* **18,** 341–359.

[4] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

[5] Cavazos-Cadena, R. and Montes-de-Oca, R. (2005). Nonstationary value iteration in controlled Markov chains with risk-sensitive average criterion. *J. Appl. Prob.* **42,** 905–918.

[6] Chávez-Rodríguez, S., Cavazos-Cadena, R. and Cruz-Suárez, H. (2016). Controlled semi-Markov chains with risk-sensitive average cost criterion. *J. Optim. Theory Appl.* **170,** 670–686.

[7] Chung, K. J. and Sobel, M. J. (1987). Discounted MDPs: distribution functions and exponential utility maximization. *SIAM J. Control Optimization* **25,** 49–62.

[8] Di Masi, G. B. and Stettner, Ł. (2007). Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. *SIAM J. Control Optimization* **46,** 231–252.

[9] Feinberg, E. A. and Rothblum, U. G. (2012). Splitting randomized stationary policies in total-reward Markov decision processes. *Math. Operat. Res.* **37,** 129–153.

[10] Ghosh, M. and Saha, S. (2014). Risk-sensitive control of continuous time Markov chains. *Stochastics* **86,** 655–675.

[11] Guo, X., Vykertas, M. and Zhang, Y. (2013). Absorbing continuous-time Markov decision processes with total cost criteria. *Adv. Appl. Prob.* **45,** 490–519.

[12] Haskell, W. B. and Jain, R. (2013). Stochastic dominance-constrained Markov decision processes. *SIAM J. Control Optimization* **51,** 273–303.

[13] Haskell, W. B. and Jain, R. (2015). A convex analytic approach to risk-aware Markov decision processes. *SIAM J. Control Optimization* **53,** 1569–1598.

[14] Hernández-Hernández, D. and Marcus, S. I. (1999). Existence of risk-sensitive optimal stationary policies for controlled Markov processes. *Appl. Math. Optimization* **40,** 273–285.

[15] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.

[16] Huang, Y. and Guo, X. (2009). Optimal risk probability for first passage models in semi-Markov decision processes. *J. Math. Anal. Appl.* **359,** 404–420.

[17] Mamer, J. W. (1986). Successive approximations for finite horizon semi-Markov decision processes with application to asset liquidation. *Operat. Res.* **34,** 638–644.

[18] 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.

[19] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.

[20] Rockafellar, R. T. (1974). *Conjugate Duality and Optimization*. SIAM, Philadelphia, PA.

[21] Ross, S. M. (1996). *Stochastic Processes*, 2nd edn. John Wiley, New York.

[22] Suresh Kumar, K. and Pal, C. (2015). Risk-sensitive ergodic control of continuous time Markov processes with denumerable state space. *Stoch. Anal. Appl.* **33,** 863–881.

https://doi.org/10.1017/apr.2018.36 Published online by Cambridge University Press