Continuous-time Markov decision processes with exponential utility

Yi Zhang [*]

Abstract: In this paper, we consider a continuous-time Markov decision process (CTMDP) in Borel spaces, where the certainty equivalent with respect to the exponential utility of the total undiscounted cost is to be minimized. The cost rate is nonnegative. We establish the optimality equation. Under the compactness-continuity condition, we show the existence of a deterministic stationary optimal policy. We reduce the risk-sensitive CTMDP problem to an equivalent risk-sensitive discrete-time Markov decision process, which is with the same state and action spaces as the original CTMDP. In particular, the value iteration algorithm for the CTMDP problem follows from this reduction. We essentially do not need impose condition on the growth of the transition and cost rate in the state, and the controlled process could be explosive.

Keywords: Continuous-time Markov decision processes. Exponential utility. Total undiscounted criteria. Risk-sensitive criterion. Optimality equation.

AMS 2000 subject classification: Primary 90C40, Secondary 60J75

1 Introduction

In this paper we consider a continuous-time Markov decision process (CTMDP) in Borel state and action spaces, where a risk averse decision maker aims at minimizing the certainty equivalent of the total undiscounted cost with respect to the exponential utility. The cost rate is nonnegative. In the literature, see e.g., [3, 5, 16, 19, 27], such a problem is traditionally also referred to as risk-sensitive, or one with exponential utility or multiplicative cost. In this paper, we use these terms interchangeably. The CTMDP with a linear utility is often called risk-neutral.

Ever since the pioneering paper [19] in 1972, there have been a large number of works devoted to risk-sensitive discrete-time Markov decision processes (DTMDPs), see [2, 4, 5, 10, 17, 21, 22, 23], to name just a few, where the term "risk-sensitive" is endowed with a more general meaning in the recent works [2, 17]. The interested reader is referred to the reference list of the aforementioned works for more relevant literature. There are some significant differences between the risk-neutral and risk-sensitive problems. For example, the criterion with exponential utility is not decomposable in the sense of [10], so that the corresponding convex analytic approach is more delicate and underdeveloped, c.f. [17]. Also, consider the model with the total discounted cost; if the state and action spaces are both finite, then there is a stationary optimal policy in the risk-neutral case, but not always in the risk-sensitive case, c.f. [21, 5]. There have also been numerous works on risk-sensitive controlled diffusions, see the reference list of [16].

In comparison, it is safe to say that much less work has been done on the risk-sensitive CTMDPs. To the best of our knowledge, the CTMDP with exponential utility was first considered in the less known article [27], where the authors considered the problem on a finite time horizon. Only verification theorems were given, i.e., the authors discussed the consequences after one has obtained a solution

[*] Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. E-mail: yi.zhang@liv.ac.uk.

to the optimality equation, provided that it satisfies certain conditions. The question of when there exits such a solution to the optimality equation was not discussed in [27]. For the same problem as in [27], this question was only considered in the recent papers [16, 32]. In [16] the transition rates were assumed to be bounded; and in [32] the growth of the transition rate was assumed to be bounded by some Lyapunov function. As a consequence, the controlled process in [16, 32] is nonexplosive under each policy. In [16, 32], the cost rate was assumed to be bounded, and the arguments are based on Dynkin's formula or the Feynman–Kac formula. The author of [32] explained why it was hard to relax this boundedness condition on the cost rate if one follows the same approach as in there, see Section 7 therein. On the other hand, one should note that unbounded transition and cost rates appear in many real-life applications; as a simplest example, consider an $M/M/\infty$ queueing system with the holding cost rate being proportional to the number of enqueued customers.

By the way, the CTMDP in [16, 27, 32] is assumed to be in a denumerable state space. In [16], the infinite horizon discounted and average problems for the risk-sensitive CTMDP were also considered.

The present paper also deals with a CTMDP with exponential utility, but is different from the aforementioned works [16, 27, 32] in the following aspects. (a) We consider the problem of minimizing the expectation of the exponential utility of the total undiscounted cost over the infinite time horizon. In the current literature, we are not aware of other work on the infinite horizon total undiscounted cost criterion for the risk-sensitive CTMDP. (b) Our method of attack does not involve the Dynkin's formula or the Feyman-Kac formula, but is based on the reduction of the risk-sensitive CTMDP to a risk-sensitive DTMDP. As an advantage of developing this approach, we basically do not need bounds on the growth of the transition and cost rates, the controlled process is allowed to be explosive, and the state space is a general Borel space. Such explosive processes are related to the "shattering into dust" phenomenon in physics, see [31]. Our reduction method starts with an adaption of Yushkevich's method [33], which was originally proposed for risk-neutral CTMDPs, and later also developed to study piecewise deterministic Markov decision processes, see [1, 7, 8]. Compared with the case of a risk-neutral CTMDP, now both the state and action spaces of the induced DTMDP are more complicated than the original CTMDP. A new equivalent DTMDP model with the same state and action spaces as the original CTMDP will be induced later after further investigations. The powerful Feinberg's reduction method for risk-neutral CTMDP, see [12], is not applicable because the criterion in the risk-sensitive CTMDP is not "decomposable".

The contributions of this paper are as follows. For a risk-sensitive CTMDP in Borel state and action spaces with the total undiscounted cost over an infinite horizon, we establish its optimality equation. Under the compactness-continuity condition, we show the existence of a deterministic stationary optimal policy. Moreover, we show that the risk-sensitive CTMDP problem is equivalent to a risk-sensitive DTMDP problem, which is with the same state and action spaces as the original CTMDP model. As a consequence of this, we also present the value iteration algorithm. Note that we only need impose rather weak conditions; the transition rate is arbitrarily unbounded, and the nonnegative cost rate is arbitrarily unbounded in the state, and the controlled process can be explosive, so that there might exist no Lyapunov function.

The rest of the paper is organize as follows. We describe the controlled process and concerned optimal control problem in Section 2. In Section 3, we present some results for the risk-sensitive DTMDPs that are needed for this paper. In Section 4, we develop Yushkevich's method to reduce the risk-sensitive CTMDP to a risk-sensitive DTMDP with more complicated state and action spaces. In Section 5, we present and prove the main results in this paper, which is ended with a conclusion in Section 6.

Notations and conventions. In what follows, $\mathcal{B}(X)$ is the Borel σ-algebra of the topological space X, I stands for the indicator function, and $\delta_{\{x\}}(\cdot)$ is the Dirac measure concentrated on the singleton $\{x\}$, assumed to be measurable. A measure is σ-additive and $[0,\infty]$-valued. Below, unless stated otherwise, the term of measurability is always understood in the Borel sense. Throughout this article, we adopt the conventions of

$$\frac{0}{0} := 0,\ 0 \cdot \infty := 0,\ \frac{1}{0} := +\infty,\ \infty - \infty := \infty. \tag{1}$$

2 Model description and problem statement

The objective of this section is to describe briefly the controlled process similarly to [12, 25, 26], and the associated optimal control problem of interest in this paper.

Let S be a nonempty Borel state space, A be a nonempty Borel action space, and q stand for a signed kernel $q(dy|x,a)$ on $\mathcal{B}(S)$ given $(x,a) \in S \times A$ such that

$$\tilde{q}(\Gamma_S|x,a) := q(\Gamma_S \setminus \{x\}|x,a) \geq 0$$

for all $\Gamma_S \in \mathcal{B}(S)$. Throughout this article we assume that $q(\cdot|x,a)$ is conservative and stable, i.e.,

$$q(S|x,a) = 0,\ \bar{q}_x = \sup_{a \in A} q_x(a) < \infty, \tag{2}$$

where $q_x(a) := -q(\{x\}|x,a)$. The signed kernel q is often called the transition rate. For simplicity and to fix ideas, we do not consider the case of different admissible action spaces at different states. Practically, the case of state-dependent admissible action spaces can be often reduced to the current setup by assigning a cost rate of ∞ at an inadmissible action, c.f. p.402 of [13].

Let us take the sample space Ω by adjoining to the countable product space $S \times ((0,\infty) \times S)^\infty$ the sequences of the form $(x_0, \theta_1, \dots, \theta_n, x_n, \infty, x_\infty, \infty, x_\infty, \dots)$, where $x_0, x_1, \dots, x_n$ belong to S, $\theta_1, \dots, \theta_n$ belong to $(0,\infty)$, and $x_\infty \notin S$ is the isolated point. We equip Ω with its Borel σ-algebra $\mathcal{F}$.

Let $t_0(\omega) := 0 =: \theta_0$, and for each $n \geq 0$, and each element $\omega := (x_0, \theta_1, x_1, \theta_2, \dots) \in \Omega$, let

$$t_n(\omega) \quad := \quad t_{n-1}(\omega) + \theta_n,$$

and

$$t_\infty(\omega) := \lim_{n \to \infty} t_n(\omega).$$

Obviously, $(t_n(\omega))$ are measurable mappings on $(\Omega, \mathcal{F})$. In what follows, we often omit the argument $\omega \in \Omega$ from the presentation for simplicity. Also, we regard x_n and θ_{n+1} as the coordinate variables, and note that the pairs $\{t_n, x_n\}$ form a marked point process with the internal history $\{\mathcal{F}_t\}_{t \geq 0}$, i.e., the filtration generated by $\{t_n, x_n\}$; see Chapter 4 of [26] for greater details. The marked point process $\{t_n, x_n\}$ defines the stochastic process $\{\xi_t, t \geq 0\}$ on $(\Omega, \mathcal{F})$ of interest by

$$\xi_t = \sum_{n \geq 0} I\{t_n \leq t < t_{n+1}\} x_n + I\{t_\infty \leq t\} x_\infty. \tag{3}$$

Here we accept $0 \cdot x := 0$ and $1 \cdot x := x$ for each $x \in S_\infty$, and below we denote $S_\infty := S \bigcup \{x_\infty\}$.

Definition 2.1 *A (history-dependent) policy π for the CTMDP is given by a sequence (π_n) such that, for each $n = 0, 1, 2, \dots$, $\pi_n(da|x_0, \theta_1, \dots, x_n, s)$ is a stochastic kernel on A, and for each $\omega = (x_0, \theta_1, x_1, \theta_2, \dots) \in \Omega$, $t > 0$,*

$$\pi(da|\omega,t) \quad = \quad I\{t \geq t_\infty\}\delta_{a_\infty}(da) + \sum_{n=0}^{\infty} I\{t_n < t \leq t_{n+1}\}\pi_n(da|x_0, \theta_1, \dots, \theta_n, x_n, t - t_n),$$

where $a_\infty \notin A$ is some isolated point. A policy $\pi = (\pi_n)$ is called Markov if, with slight abuse of notations, each of the stochastic kernels π_n reads $\pi_n(da|x_0, \theta_1, \ldots, x_n, s) = \pi_n(da|x_n, s)$. A Markov policy is further called deterministic if the stochastic kernels $\pi_n(da|x_n, s)$ all degenerate. A policy $\pi = (\pi_n)$ is called stationary if, with slight abuse of notations, each of the stochastic kernels π_n reads $\pi_n(da|x_0, \theta_1, \ldots, x_n, s) = \pi(da|x_n)$. A stationary policy is further called deterministic if $\pi_n(da|x_0, \theta_1, \ldots, x_n, s) = \delta_{\{f(x_n)\}}(da)$ for some measurable mapping f from S to A. We shall identify such a deterministic stationary policy by the underlying measurable mapping f.

The class of all policies for the CTMDP model is denoted by Π.

Under a policy $\pi := (\pi_n) \in \Pi$, we define the following random measure on $S \times (0, \infty)$

$$\begin{aligned}\nu^\pi(dy, dt) &:= \int_A \tilde{q}(dy|\xi_{t-}(\omega), a)\pi(da|\omega, t)dt \\ &= \sum_{n\geq 0} \int_A \tilde{q}(dy|x_n, a)\pi_n(da|x_0, \theta_1, \ldots, \theta_n, x_n, t - t_n)I\{t_n < t \leq t_{n+1}\}dt\end{aligned}$$

with $q_{x_\infty}(a_\infty) = q(dy|x_\infty, a_\infty) := 0 =: q_{x_\infty}(a)$ for each $a \in A$. Recall that ω is the generic notation of an element of Ω. Then for each given initial distribution γ on $\mathcal{B}(S)$, there exists a unique probability measure P^π_γ such that

$$P^\pi_\gamma(x_0 \in dx) = \gamma(dx),$$

and with respect to P^π_γ, ν^π is the dual predictable projection of the random measure associated with the marked point process $\{t_n, x_n\}$; see [20, 26]. The process $\{\xi_t\}$ defined by (3) under the probability measure P^π_γ is called a CTMDP. Below, when γ is a Dirac measure concentrated at $x \in S$, we use the denotation P^π_x. Expectations with respect to P^π_γ and P^π_x are denoted as E^π_γ and E^π_x, respectively.

The following remark follows from [20].

Remark 2.1 *Under a fixed policy $\pi = (\pi_n)$, the conditional distribution of (θ_{n+1}, x_{n+1}) with the condition on $x_0, \theta_1, x_1, \ldots, \theta_n, x_n$ is given on $\{\omega : x_n(\omega) \in S\}$ by*

$$\begin{aligned}& P^\pi_\gamma(\theta_{n+1} \in \Gamma_1,\ x_{n+1} \in \Gamma_2|x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \\ = & \int_{\Gamma_1} e^{-\int_0^t \int_A q_{x_n}(a)\pi_n(da|x_0,\theta_1,\ldots,\theta_n,x_n,s)ds} \int_A \tilde{q}(\Gamma_2|x_n, a)\pi_n(da|x_0, \theta_1, \ldots, \theta_n, x_n, t)dt, \\ & \forall\ \Gamma_1 \in \mathcal{B}((0, \infty)),\ \Gamma_2 \in \mathcal{B}(S); \\ & P^\pi_\gamma(\theta_{n+1} = \infty,\ x_{n+1} = x_\infty|x_0, \theta_1, x_1, \ldots, \theta_n, x_n) = e^{-\int_0^\infty \int_A q_{x_n}(a)\pi_n(da|x_0,\theta_1,\ldots,\theta_n,x_n,s)ds},\end{aligned}$$

and given on $\{\omega : x_n(\omega) = x_\infty\}$ by

$$P^\pi_\gamma(\theta_{n+1} = \infty,\ x_{n+1} = x_\infty|x_0, \theta_1, x_1, \ldots, \theta_n, x_n) = 1.$$

Let the cost rate be given by a $[0, \infty)$-valued measurable function c on $S \times A$. In this paper, we study the following optimal control problem:

$$\text{Minimize over } \pi \in \Pi: \qquad E^\pi_x\left[e^{\int_0^\infty \int_A c(\xi_t,a)\pi(da|\omega,t)dt}\right] =: V(x, \pi),\ x \in S. \tag{4}$$

Here and below, we put $c(x_\infty, a) := 0$ for each $a \in A$.

The CTMDP problem (4) is equivalent to minimizing the certainty equivalent of the total cost with respect to the exponential utility for a risk averse decision maker, see [4, 16, 19].

In what follows, we refer the CTMDP problem (4) with the exponential utility to as the CTMDP model $\{S, A, q, c\}$.

Definition 2.2 *A policy π^* is called optimal for problem (4) if*

$$V(x,\pi^*) = \inf_{\pi\in\Pi} V(x,\pi) =: V^*(x),\ \forall\ x \in S.$$

Evidently, $V^*(x) \geq 1$ for each $x \in S$.

One powerful method of reducing a CTMDP to a DTMDP is due to Yushkevich [33], which considers the case of a linear utility. However, the induced DTMDP is with a more complicated action space, so that a deterministic stationary strategy in the induced DTMDP in general does not give a deterministic stationary policy for the CTMDP model, but gives a specific Markov policy. This approach has been further developed to study piecewise deterministic Markov decision processes, see the books [1, 7, 8]. In Section 4, as a preparation for our main optimality result, we shall develop this method for the case of exponential utility. In contrast to the linear utility case, now both the state and action spaces of the reduced DTMDP are more complicated than those of the CTMDP; e.g., a deterministic stationary strategy in this reduced DTMDP will no longer give a Markov policy for the CTMDP model. A further reduction to a simpler DTMDP with the same state and action spaces as the original CTMDP will be given in a subsequent section.

3 Discrete-time Markov decision process with exponential utility

To serve the investigations of the CTMDP, in this section we present briefly the dynamic programming approach for the DTMDP model (with exponential utility). The presented results are mostly related to [23], which however, is based on the compactness-continuity condition. For our purpose, we would not assume any compactness-continuity condition here, except for Proposition 3.4, and would need to consider a slightly more general cost function, as compared to [23]. Without assuming the compactness-continuity condition, the dynamic programming approach for the DTMDP model was partially studied in [3], which dealt with a bounded cost function and mainly a finite horizon, see p.90 and Section 11.3 therein. Instead of the dynamic programming approach, [10, 11] developed a different method for studying a rather general class of DTMDP problems in Borel spaces. That method, which can be traced back to Girsanov, is based on the investigations of strategic measures, and does not give all the results we would need here.

Consider a discrete-time Markov decision process with the following primitives:

- $\mathbf{X}$ is a nonempty Borel state space.
- $\mathbf{A}$ is a nonempty Borel action space.
- $p(dy|x,a)$ is a stochastic kernel on $\mathcal{B}(\mathbf{X})$ given $(x,a) \in \mathbf{X}\times\mathbf{A}$.
- l a $[0,\infty]$-valued measurable cost function on $\mathbf{X}\times\mathbf{A}\times\mathbf{X}$.

Let us denote for each $n = 1,2,\dots,\infty$, $\mathbf{H}_n := \mathbf{X}\times(\mathbf{A}\times\mathbf{X})^n$ and $\mathbf{H}_0 := \mathbf{X}$. A strategy $\sigma = (\sigma_n)_{n=0}^{\infty}$ in the DTMDP is given by a sequence of stochastic kernels $\sigma_n(da|h_n)$ on $\mathcal{B}(\mathbf{A})$ from $h_n \in \mathbf{H}_n$ for $n = 0,1,2,\dots$. A strategy $\sigma = (\sigma_n)$ is called deterministic Markov if for each $n = 0,1,2,\dots$, $\sigma_n(da|h_n) = \delta_{\{\varphi_n(x_n)\}}(da)$, where φ_n is an $\mathbf{A}$-valued measurable mapping on $\mathbf{X}$. We identify such a deterministic Markov strategy with (φ_n).

Let Σ be the space of strategies, and Σ_{DM} be the space of all deterministic strategies for the DTMDP.

Let the controlled and controlling processes be denoted by $\{Y_n, n = 0,1,\dots,\infty\}$ and $\{A_n, n = 0,1,\dots,\infty\}$, respectively. Here, for each $n = 0,1,\dots$, Y_n is the projection of $\mathbf{H}_\infty$ to the $2n+1$st coordinate, and A_n to the $2n+2$nd coordinate.

Under a strategy $\sigma = (\sigma_n)$ and a given initial probability distribution ν on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$, by the Ionescu-Tulcea theorem, c.f., [18, 28], there is a probability measure $\mathbf{P}_\nu^\sigma$ on $(\mathbf{H}_\infty, \mathcal{B}(\mathbf{H}_\infty))$ such that

$$\begin{aligned}
&\mathbf{P}_\nu^\sigma(Y_0 \in dx) = \nu(dx), \\
&\mathbf{P}_\nu^\sigma(A_n \in da|Y_0, A_0, \dots, Y_n) = \sigma_n(da|Y_0, A_0, \dots, Y_n),\ n = 0, 1, \dots, \\
&\mathbf{P}_\nu^\sigma(Y_{n+1} \in dx|Y_0, A_0, \dots, Y_n, A_n) = p(dx|Y_n, A_n),\ n = 0, 1, \dots.
\end{aligned}$$

As usual, equalities involving conditional expectations and probabilities are understood in the almost sure sense. The probability measure $\mathbf{P}_\nu^\sigma$ is called a strategic measure for the DTMDP. The expectation taken with respect to $\mathbf{P}_\nu^\sigma$ is denoted by $\mathbf{E}_\nu^\sigma$. When ν is concentrated on the singleton $\{x\}$, $\mathbf{P}_\nu^\sigma$ and $\mathbf{E}_\nu^\sigma$ are written as $\mathbf{P}_x^\sigma$ and $\mathbf{E}_x^\sigma$.

Consider the optimal control problem

$$\text{Minimize over } \sigma : \qquad \mathbf{E}_x^\sigma\left[e^{\sum_{n=0}^\infty l(Y_n, A_n, Y_{n+1})}\right] =: \mathbf{V}(x, \sigma),\ x \in \mathbf{X}. \tag{5}$$

We denote the value function of problem (5) by $\mathbf{V}^*$. Then a strategy σ^* is called optimal for problem (5) if $\mathbf{V}(x, \sigma^*) = \mathbf{V}^*(x)$ for each $x \in \mathbf{X}$. We refer problem (5) to as the DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ (with the exponential utility).

Note that

$$\mathbf{V}^*(x) \geq 1,\ \forall\ x \in \mathbf{X}. \tag{6}$$

One can write

$$\mathbf{E}_x^\sigma\left[e^{\sum_{n=0}^\infty l(Y_n, A_n, Y_{n+1})}\right] = \int_{\mathbf{H}_\infty} e^{\sum_{n=0}^\infty l(Y_n, A_n, Y_{n+1})} \mathbf{P}_x^\sigma(dh),\ x \in \mathbf{X}. \tag{7}$$

Then $\mathbf{V}(x, \cdot)$ is a measurable criterion in the sense of [10, 11]. In view of this, that $\mathbf{V}^*$ is a lower semianalytic function on $\mathbf{X}$ immediately follows from Theorem 4.2 of [11].

Proposition 3.1 *The function $\boldsymbol{V}^*$ is an $[1, \infty]$-valued lower semianalytic solution to*

$$\boldsymbol{V}(x) = \inf_{a \in \boldsymbol{A}} \left\{ \int_{\boldsymbol{X}} p(dy|x, a) e^{l(x,a,y)}\, \boldsymbol{V}(y) \right\},\ x \in \boldsymbol{X}. \tag{8}$$

Proof. Let σ be an arbitrarily fixed strategy. For each $x \in \mathbf{X}$ and $b \in \mathbf{A}$, consider the shifted strategy $\sigma^{(x,b)} = (\sigma_n^{(x,b)})$ by $\sigma_n^{(x,b)}(da|h_n) = \sigma_{n+1}(da|x, b, h_n)$ for each $n = 0, 1, \dots,$ and $h_n \in \mathbf{H}_n$. Then for each fixed $x \in \mathbf{X}$,

$$\mathbf{V}^*(x) \leq \mathbf{V}(x, \sigma^{(x,b)}) = \mathbf{E}_x^\sigma\left[e^{\sum_{n=1}^\infty l(Y_n, A_n, Y_{n+1})}\middle| Y_0 = x, A_0 = b, Y_1 = y\right],$$

where the equality holds almost surely with respect to $p(dy|x, b)\sigma_0(db|x)$. Now for each $x \in \mathbf{X}$,

$$\begin{aligned}
\mathbf{V}(x, \sigma) &= \mathbf{E}_x^\sigma\left[e^{l(Y_0, A_0, Y_1)} e^{\sum_{n=1}^\infty l(Y_n, A_n, Y_{n+1})}\right] \\
&= \mathbf{E}_x^\sigma\left[e^{l(Y_0, A_0, Y_1)} \mathbf{E}_x^\sigma\left[e^{\sum_{n=1}^\infty l(Y_n, A_n, Y_{n+1})}\middle| Y_0, A_0, Y_1\right]\right] \\
&= \int_{\mathbf{A}} \int_{\mathbf{X}} p(dy|x, b) e^{l(x,b,y)} \mathbf{E}_x^\sigma\left[e^{\sum_{n=1}^\infty l(Y_n, A_n, Y_{n+1})}\middle| Y_0 = x, A_0 = b, Y_1 = y\right] \sigma_0(db|x) \\
&\geq \int_{\mathbf{A}} \int_{\mathbf{X}} p(dy|x, b) e^{l(x,b,y)} \mathbf{V}^*(y) \sigma_0(db|x) \\
&\geq \inf_{a \in \mathbf{A}} \left\{ \int_{\mathbf{X}} p(dy|x, a) e^{l(x,a,y)} \mathbf{V}^*(y) \right\},
\end{aligned}$$

where and below integrals such as those in the above inequalities are well defined because $\mathbf{V}^*$ is lower semianalytic, see Lemma 7.30 as well as Proposition 7.48 of [3]. Thus,

$$\mathbf{V}^*(x) \geq \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{X}} p(dy|x,a) e^{l(x,a,y)} \mathbf{V}^*(y) \right\}, \ x \in \mathbf{X}. \tag{9}$$

Next, we establish the opposite direction of the above inequality. Let $z \in \mathbf{X}$, $b \in \mathbf{A}$ and $\epsilon > 0$ be arbitrarily fixed. It follows from Theorem 3.1 of [10] (c.f. Chapter 3 of [9]) that there exists a strategy $\hat{\sigma}$ such that

$$\ln \mathbf{V}^*(x) + \epsilon \geq \ln \mathbf{E}_x^{\hat{\sigma}} \left[e^{\sum_{n=0}^{\infty} l(Y_n,A_n,Y_{n+1})} \right] = \ln \mathbf{V}(x,\hat{\sigma})$$

for almost all $x \in \mathbf{X}$ with respect to $p(dx|z,b)$. Consider the strategy $\sigma' = (\delta_{\{b\}}, \hat{\sigma}_0, \hat{\sigma}_1, \dots)$. Then

$$\begin{aligned} \mathbf{V}^*(z) &\leq \mathbf{E}_z^{\sigma'} \left[e^{\sum_{n=0}^{\infty} l(Y_n,A_n,Y_{n+1})} \right] = \int_{\mathbf{X}} p(dy|z,b) e^{l(z,b,y)} e^{\ln \mathbf{V}(y,\hat{\sigma})} \\ &\leq \int_{\mathbf{X}} p(dy|z,b) e^{l(z,b,y)} e^{\ln \mathbf{V}^*(y)+\epsilon} = e^{\epsilon} \int_{\mathbf{X}} p(dy|z,b) e^{l(z,b,y)} \mathbf{V}^*(y). \end{aligned}$$

Since $\epsilon > 0$, $z \in \mathbf{X}$ and $b \in \mathbf{A}$ are arbitrarily fixed, it follows from the above that

$$\mathbf{V}^*(x) \leq \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{X}} p(dy|x,a) e^{l(x,a,y)} \mathbf{V}^*(y) \right\}, \ x \in \mathbf{X}.$$

Combing this with (9), we see that the statement holds. □

Proposition 3.2 *For each $x \in \boldsymbol{X}$,*

$$\inf_{\sigma\in\Sigma_{DM}} \boldsymbol{V}(x,\sigma) = \inf_{\sigma\in\Sigma} \boldsymbol{V}(x,\sigma),$$

where we recall that Σ_{DM} is the set of all deterministic Markov strategies for the DTMDP.

The proof of this proposition is based on the next result.

Lemma 3.1 *Let f be a $[1,\infty]$-valued lower semianalytic function on the Borel space $\boldsymbol{X} \times \boldsymbol{A}$, and f^* be a function on $\boldsymbol{X}$ defined by $f^*(x) = \inf_{a\in\boldsymbol{A}} f(x,a)$ for each $x \in \boldsymbol{X}$. Then for each $\epsilon > 0$, there exists an analytically measurable mapping φ from $\boldsymbol{X}$ to $\boldsymbol{A}$ such that*

$$f(x,\varphi(x)) \leq f^*(x)e^{\epsilon}, \ \forall \ x \in \boldsymbol{X}.$$

Proof. The reasoning of Proposition 7.50 of [3] can be easily modified to prove the statement of this lemma. The details are omitted. □

Now we are in position to prove Proposition 3.2.

Proof of Proposition 3.2. Let $\epsilon > 0$ and $x_0 \in \mathbf{X}$ be arbitrarily fixed. We first show that there exists a deterministic Markov strategy σ such that

$$\mathbf{V}(x_0,\sigma) \leq \mathbf{V}^*(x_0)e^{\epsilon} \tag{10}$$

as follows.

Let (ϵ_k) be a sequence of positive constants such that $\sum_{k=0}^{\infty}\epsilon_k=\epsilon$. By Proposition 3.1, there is a Borel measurable mapping φ_0 from $\mathbf{X}$ to $\mathbf{A}$ such that

$$\mathbf{V}^*(x_0)\geq e^{-\epsilon_0}\int_{\mathbf{X}}p(dx_1|x_0,\varphi_0(x_0))e^{l(x_0,\varphi_0(x_0),x_1)}\mathbf{V}^*(x_1). \tag{11}$$

(Remember, the above inequality is only required to hold for the fixed $x_0\in\mathbf{X}$.) By Lemma 3.1 and Proposition 3.1, for each $k=1,2,\dots,$ there exists an analytically measurable mapping $\tilde{\varphi}_k$ from $\mathbf{X}$ to $\mathbf{A}$ such that

$$\mathbf{V}^*(x)\geq e^{-\epsilon_k}\int_{\mathbf{X}}p(dy|x,\tilde{\varphi}_k(x))e^{l(x,\tilde{\varphi}_k(x),y)}\mathbf{V}^*(y),\ \forall\ x\in\mathbf{X}. \tag{12}$$

Let φ_1 be the Borel measurable modification of $\tilde{\varphi}_1$ with respect to the probability measure $p(\cdot|x_0,\varphi_0(x_0))$. Then

$$\begin{aligned}\mathbf{V}^*(x_0)&\geq e^{-\epsilon_0-\epsilon_1}\int_{\mathbf{X}}p(dx_1|x_0,\varphi_0(x_0))e^{l(x_0,\varphi_0(x_0),x_1)}\int_{\mathbf{X}}p(dx_2|x_1,\tilde{\varphi}_1(x_1))e^{l(x_1,\tilde{\varphi}_1(x_1),x_2)}\mathbf{V}^*(x_2)\\&\geq e^{-\epsilon_0-\epsilon_1}\int_{\mathbf{X}}p(dx_1|x_0,\varphi_0(x_0))e^{l(x_0,\varphi_0(x_0),x_1)}\int_{\mathbf{X}}p(dx_2|x_1,\varphi_1(x_1))e^{l(x_1,\varphi_1(x_1),x_2)},\end{aligned}$$

where the first inequality is by (11) and (12), and the second inequality is by (6). Inductively, for each $k=2,3,\dots,$ let φ_k be a Borel measurable modification of $\tilde{\varphi}_k$ with respect to the probability measure $\int_{\mathbf{X}^{k-1}}p(\cdot|x_{k-1},\varphi_{k-1}(x_{k-1}))p(dx_{k-1}|x_{k-2},\varphi_{k-2}(x_{k-2}))\dots p(dx_1|x_0,\varphi_0(x_0))$. Let $\sigma=(\varphi_n)_{n=0}^{\infty}$ be the deterministic Markov strategy. Then

$$\mathbf{V}^*(x_0)\ \geq\ e^{-\sum_{n=0}^{k}\epsilon_n}\mathbf{E}_{x_0}^{\sigma}\left[e^{\sum_{n=0}^{k}l(Y_n,A_n,Y_{n+1})}\right],\ \forall\ k=0,1,\dots.$$

By passing to the limit as $k\to\infty$ in the above inequality, we see that (10) holds.

Now the statement of the theorem follows from (10) and the arbitrariness of $\epsilon>0$ and $x_0\in\mathbf{X}$. □

Proposition 3.3 *The following two assertions hold.*

(a) Let $\boldsymbol{U}$ be a $[1,\infty]$-valued lower semianalytic function on $\boldsymbol{X}$. If

$$\boldsymbol{U}(x)\geq\inf_{a\in\boldsymbol{A}}\left\{\int_{\boldsymbol{X}}p(dy|x,a)e^{l(x,a,y)}\,\boldsymbol{U}(y)\right\},\ \forall\ x\in\boldsymbol{X},$$

then $\boldsymbol{U}(x)\geq\boldsymbol{V}^*(x)$ for each $x\in\boldsymbol{X}$. In particular, if the function $\boldsymbol{U}$ satisfying the above relation is $[1,\infty)$-valued, then so is the value function $\boldsymbol{V}^*$.

(b) Let φ be a deterministic stationary strategy for the DTMDP model $\{\boldsymbol{X},\boldsymbol{A},p,l\}$. If

$$\boldsymbol{V}^*(x)=\int_{\boldsymbol{X}}p(dy|x,\varphi(x))e^{l(x,\varphi(x),y)}\,\boldsymbol{V}^*(y),\ \forall\ x\in\boldsymbol{X}, \tag{13}$$

then $\boldsymbol{V}^*(x)=\boldsymbol{V}(x,\varphi)$ for each $x\in\boldsymbol{X}$.

Proof. (a) Let $x\in\mathbf{X}$ and $\epsilon>0$ be fixed. Then by using Lemma 3.1, one can follow the reasoning in the proof of Proposition 3.2, and see the existence of a deterministic Markov strategy σ, which satisfies

$$\mathbf{U}(x)\geq e^{-\epsilon}\mathbf{V}(x,\sigma)\geq e^{-\epsilon}\mathbf{V}^*(x).$$

Since $\epsilon > 0$ and $x \in \mathbf{X}$ are arbitrarily fixed, the statement follows.

(b) Consider the given deterministic stationary strategy φ. Let $x \in \mathbf{X}$ be fixed. Then by iterations based on (13), and keeping in mind that $\mathbf{V}^*$ is $[1, \infty]$-valued, we see

$$\mathbf{V}^*(x) = \mathbf{E}_x^\varphi \left[e^{\sum_{n=0}^m l(Y_n, A_n, Y_{n+1})} \mathbf{V}(Y_{m+1}) \right] \geq \mathbf{E}_x^\varphi \left[e^{\sum_{n=0}^m l(Y_n, A_n, Y_{n+1})} \right]$$

for each $m = 1, 2, \ldots$. Thus, the statement holds after passing to the limit as $m \to \infty$. □

Condition 3.1 *(a) The function l is lower semicontinuous on $\boldsymbol{X} \times \boldsymbol{A} \times \boldsymbol{X}$.*

(b) For each bounded continuous function f on $\boldsymbol{X}$, $\int_{\boldsymbol{X}} f(y)p(dy|x,a)$ is continuous in $(x,a) \in \boldsymbol{X} \times \boldsymbol{A}$.

(c) The space $\boldsymbol{A}$ is a compact Borel space.

Condition 3.2 *(a) The function $l(x,a,y)$ is lower semicontinuous in $a \in \boldsymbol{A}$ for each $x, y \in \boldsymbol{X}$.*

(b) For each bounded measurable function f on $\boldsymbol{X}$ and each $x \in \boldsymbol{X}$, $\int_{\boldsymbol{X}} f(y)p(dy|x,a)$ is continuous in $a \in \boldsymbol{A}$.

(c) The space $\boldsymbol{A}$ is a compact Borel space.

Proposition 3.4 *(a) Suppose Condition 3.1 is satisfied. Then the value function $\boldsymbol{V}^*$ is the minimal $[1, \infty]$-valued lower semicontinuous solution to (8).*

(b) Suppose Condition 3.2 is satisfied. Then the value function $\boldsymbol{V}^*$ is the minimal $[1, \infty]$-valued measurable solution to (8).

(c) Suppose Condition 3.1 or Condition 3.2 is satisfied. Let $\boldsymbol{V}^{(0)}(x) := 1$ for each $x \in \boldsymbol{X}$, and for each $n = 1, 2, \ldots$,

$$\boldsymbol{V}^{(n)}(x) := \inf_{a \in A} \left\{ \int_{\boldsymbol{X}} p(dy|x,a) e^{l(x,a,y)} \, \boldsymbol{V}^{(n-1)}(y) \right\}, \ \forall \ x \in \boldsymbol{X}.$$

Then $(\boldsymbol{V}^{(n)}(x))$ increases to $\boldsymbol{V}^*(x)$ for each $x \in \boldsymbol{X}$, where $\boldsymbol{V}^*$ is the value function for problem (5). Furthermore, there exists a deterministic stationary strategy φ satisfying (13), and so in particular, there exists a deterministic stationary optimal strategy for the DTMDP problem (5).

Proof. This statement can be proved as in [23]. □

4 First reduction to a DTMDP model

In this section, we reduce the risk-sensitive CTMDP to a risk-sensitive DTMDP with more complicated state and action spaces. As in [33], this is based on viewing a policy $\pi = (\pi_n)$ as a sequence of measurable mappings taking values in the (quotient) space of $\mathbb{P}(A)$-valued measurable mappings, where and below, the set $\mathbb{P}(A)$ is the space of probability measures on $\mathcal{B}(A)$, and is equipped with the standard weak topology, so that $\mathbb{P}(A)$ is a Borel space, see Chapter 7 of [3]. The details are as follows.

Let $\mathcal{R}$ denote the set of (Borel) measurable mappings $\rho_t(da)$ from $t \in (0, \infty) \to \mathbb{P}(A)$. Here, we do not distinguish between two measurable mappings in $t \in (0, \infty)$ which coincide almost everywhere with respect to the Lebesgue measure.

We endow $\mathcal{R}$ with the σ-algebra as the minimal one with respect to which, the function

$$\rho \in \mathcal{R} \to \int_0^\infty e^{-t} f(t, \rho_t) dt$$

is measurable in $\rho \in \mathcal{R}$ for each bounded measurable function f on $(0,\infty) \times \mathbb{P}(A)$. Lemma 1 of [33] asserts that $\mathcal{R}$ is a Borel space.

For the rest of this paper, it is convenient to introduce the following notations. For each $\mu \in \mathbb{P}(A)$,

$$\begin{aligned} q_x(\mu) &:= \int_A q_x(a)\mu(da); \\ \tilde{q}(dy|x,\mu) &:= \int_A \tilde{q}(dy|x,a)\mu(da); \\ c(x,\mu) &:= \int_A c(x,a)\mu(da). \end{aligned}$$

It follows from [33] that we can legitimately consider a DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ with exponential utility with the following primitives, where all the functions and mappings are measurable.

- The state space is $\mathbf{X} := ((0,\infty) \times S) \bigcup \{(\infty, x_\infty)\}$. Whenever the topology is concerned, (∞, x_∞) is regarded as an isolated point in $\mathbf{X}$.
- The action space is $\mathbf{A} := \mathcal{R}$.
- The transition kernel p on $\mathcal{B}(\mathbf{X})$ from $\mathbf{X} \times \mathbf{A}$ is given for each $\rho \in \mathbf{A}$ by

$$\begin{aligned} p(\Gamma_1 \times \Gamma_2|(\theta,x),\rho) &:= \int_{\Gamma_2} e^{-\int_0^t q_x(\rho_s)ds}\tilde{q}(\Gamma_1|x,\rho_t)dt, \\ & \quad\ \forall\ \Gamma_1 \in \mathcal{B}(S),\ \Gamma_2 \in \mathcal{B}((0,\infty)),\ x \in S,\ \theta \in (0,\infty), \\ p(\{\infty\} \times \{x_\infty\}|(\theta,x),\rho) &:= e^{-\int_0^\infty q_x(\rho_s)ds},\ \forall\ x \in S,\ \theta \in (0,\infty); \\ p(\{(\infty,x_\infty)\}|(\infty,x_\infty),\rho) &:= 1. \end{aligned} \tag{14}$$

- The cost function l is a $[0,\infty]$-valued measurable function on $\mathbf{X} \times \mathbf{A} \times \mathbf{X}$ given by

$$l((\theta,x),\rho,(\tau,y)) := \int_0^\infty I\{s<\tau\}c(x,\rho_s)ds =: \hat{l}(x,\rho,\tau),\ \forall\ ((\theta,x),\rho,(\tau,y)) \in \mathbf{X} \times \mathbf{A} \times \mathbf{X}. \tag{15}$$

(Recall that $c(x_\infty,a) := 0$ and $q_{x_\infty}(a) := 0 =: \tilde{q}(S|x_\infty,a)$ for each $a \in A$.) For each strategy σ for the DTMDP $\{\mathbf{X}, \mathbf{A}, p, l\}$, the function $\mathbf{V}(\cdot,\sigma)$ is defined by (5).

The controlled process in the above DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ is denoted by $\{Y_n, n = 0, 1, \dots\}$, where $Y_n = (\Theta_n, X_n)$, and the controlling process is denoted by $\{A_n, n = 0, 1, \dots\}$. For $n \geq 1$, Θ_n and X_n correspond to the nth sojourn time and the post-jump state in the CTMDP, Θ_0 is fictitious, and X_0 is the initial state in the CTMDP. Let Σ^0_{DM} be the class of deterministic Markov strategies for the DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ in the form $\sigma = (\varphi_n)$ where $\varphi_0((\theta,x))$ does not depend on $\theta \in (0,\infty)$ for each $x \in S$.

For each fixed $\hat{\theta} \in (0,\infty)$ and a deterministic Markov strategy $\sigma = (\varphi_n)$,

$$\mathbf{V}((\hat{\theta},x),\sigma) = \mathbf{V}((\hat{\theta},x),\hat{\sigma}) = \mathbf{V}((\theta,x),\hat{\sigma}),\ \forall\ x \in S,\ \theta \in (0,\infty), \tag{16}$$

where $\hat{\sigma} = (\hat{\varphi}_0, \varphi_1, \varphi_2, \dots)$ with $\hat{\varphi}_0((\theta,x)) = \varphi_0((\hat{\theta},x))$ for each $\theta \in (0,\infty)$. Indeed, since $\hat{\theta}$ is fixed, under the strategy $\hat{\sigma}$, the decision is made independently of the first coordinate of the initial state. This, together with the definitions of the transition kernel p and the cost function l given by (14) and (15), justifies (16). See also Theorem 2 of [13].

Proposition 3.2 and (16) imply that

$$\mathbf{V}^*((\theta,x)) = \inf_{\sigma\in\Sigma^0_{DM}} \mathbf{V}((\theta,x),\sigma),\ \forall\ x\in S,\ \theta\in(0,\infty),$$

where $\mathbf{V}^*$ is the value function of the DTMDP problem (5). Since for each $\sigma\in\Sigma^0_{DM}$, $V(\theta,x,\sigma) = V(\theta',x,\sigma)$ for each $\theta,\theta'\in(0,\infty)$, it follows that $\mathbf{V}^*((\theta,x))$ does not depend on $\theta\in(0,\infty)$. Therefore, we write $\mathbf{V}^*(x)$ instead of $\mathbf{V}^*((\theta,x))$ and $\mathbf{V}(x,\sigma)$ instead of $\mathbf{V}((\theta,x),\sigma)$ when σ is in Σ^0_{DM}. The previous equality now reads

$$\mathbf{V}^*(x) = \inf_{\sigma\in\Sigma^0_{DM}} \mathbf{V}(x,\sigma),\ \forall\ x\in S. \tag{17}$$

Consider a policy $\pi=(\pi_n)$ for the CTMDP model $\{S,A,q,c\}$. Note that each stochastic kernel

$$\pi_n(da|x_0,\theta_1,x_1,\theta_2,\dots,\theta_n,x_n,s)$$

can be identified with a measurable mapping say $\pi_n(x_0,\theta_1,x_1,\theta_2,\dots,\theta_n,x_n)(s,da)$ from $S\times\mathbf{X}^n$ to $\mathcal{R}$, and vice versa, see Lemma 3 of [33]. Therefore, each policy $\pi=(\pi_n)$ for the CTMDP model $\{S,A,q,c\}$ is identified with a deterministic strategy denoted by $\sigma(\pi)$ for the DTMDP model $\{\mathbf{X},\mathbf{A},p,l\}$, where, under this strategy $\sigma(\pi)$, at the time step n, the decision in $\mathbf{A}$ is made based only on $X_0,Y_1,Y_2,\dots,Y_n$ and n, and is independent on Θ_0 and the past actions. Therefore, under the policy $\pi=(\pi_n)$ for the CTMDP model $\{S,A,q,c\}$, for each $x\in S$ and $\theta\in(0,\infty)$,

$$\begin{aligned} V(x,\pi) &= E_x^\pi\left[e^{\int_0^\infty\int_A c(x,a)\pi(da|\omega,t)dt}\right] = E_x^\pi\left[e^{\sum_{n=0}^\infty\int_0^{\theta_{n+1}}\int_A c(x_n,a)\pi_n(da|x_0,\theta_1,\dots,\theta_n,x_n,s)ds}\right] \\ &= \mathbf{E}^{\sigma(\pi)}_{(\theta,x)}\left[e^{\sum_{n=0}^\infty l(Y_n,A_n,Y_{n+1})}\right] = \mathbf{E}^{\sigma(\pi)}_{(\theta,x)}\left[e^{\sum_{n=0}^\infty \hat{l}(X_n,A_n,\Theta_{n+1})}\right] = \mathbf{V}(x,\sigma(\pi)), \end{aligned}$$

c.f. Remark 2.1, (14) and (15) for the third equality, and Theorem 2 of [13] for the last equality. Thus,

$$V^*(x)\geq\mathbf{V}^*(x),\ \forall\ x\in S.$$

On the other hand, each deterministic Markov strategy $\sigma=(\varphi_n)\in\Sigma^0_{DM}$ can be identified with a policy say $\pi(\sigma)$ for the CTMDP model $\{S,A,q,c\}$ such that

$$V(x,\pi(\sigma)) = \mathbf{V}(x,\sigma),\ \forall\ x\in S.$$

This and (17) imply

$$V^*(x)\leq\mathbf{V}^*(x),\ \forall\ x\in S.$$

Now we come to the following conclusion.

Theorem 4.1 *The value function V^* for the CTMDP problem (4) is lower semianalytic on S and satisfies*

$$V^*(x) = \boldsymbol{V}^*(x),\ \forall\ x\in S.$$

Furthermore, V^* satisfies

$$\begin{aligned} V^*(x) &= \inf_{\rho\in\mathcal{R}}\left\{\int_0^\infty e^{-\int_0^\tau(q_x(\rho_s)-c(x,\rho_s))ds}\left(\int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)\right)d\tau + e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c(x,\rho_s)ds}\right\}, \\ &\quad \forall\ x\in S. \end{aligned} \tag{18}$$

Proof. The equality between V^* and $\mathbf{V}^*$ on S follows from the discussions above the theorem. The other parts of the statement are then by Proposition 3.1. □

In (18), it is possible that $\int_0^\infty q_x(\rho_s)ds = \infty = \int_0^\infty c(x,\rho_s)ds$, so that by (1)

$$e^{-\int_0^\infty q_x(\rho_s)ds} e^{\int_0^\infty c(x,\rho_s)ds} = 0\cdot\infty = 0 \neq \infty = e^{\infty} = e^{-\infty+\infty} = e^{-\int_0^\infty q_x(\rho_s)ds + \int_0^\infty c(x,\rho_s)ds}.$$

On the other hand, by (2),

$$e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} = e^{-\int_0^\tau q_x(\rho_s)ds} e^{\int_0^\tau c(x,\rho_s)ds}$$

for each $\tau\in(0,\infty)$.

5 Optimality result

In this section, we establish the optimality equation for the CTMDP problem (4); show, under some compactness-continuity conditions, the existence of a deterministic stationary optimal policy for problem (4); and further reduce the CTMDP model $\{S,A,q,c\}$ to a simpler DTMDP model with the same state and action spaces as the CTMDP, in contrast to the DTMDP model in Section 4. As a corollary of this reduction, we formulate the value iteration algorithm for problem (4).

5.1 Optimality equation

In this subsection, we establish the optimality equation satisfied by the value function V^* of the CTMDP problem (4). This is done based on more detailed investigations of (18).

Theorem 5.1 *(a) The value function V^* of the CTMDP problem (4) is an $[1,\infty]$-valued lower semianalytic function satisfying*

$$0 = \inf_{a\in A}\left\{c(x,a)V^*(x) + \int_S q(dy|x,a)V^*(y)\right\} \tag{19}$$

for each $x\in S$ such that $V^*(x)<\infty$.

(b) If a deterministic stationary policy φ for the CTMDP model $\{S,A,q,c\}$ satisfies

$$0 = \inf_{a\in A}\left\{c(x,a)V^*(x) + \int_S q(dy|x,a)V^*(y)\right\} = c(x,\varphi(x))V^*(x) + \int_S q(dy|x,\varphi(x))V^*(y) \tag{20}$$

for each $x\in S$ such that $V^*(x)<\infty$, then the deterministic stationary policy φ is optimal for the CTMDP problem (4). (The definition of φ on $\{x\in S:\ V^*(x)=\infty\}$ can be put arbitrarily, so long φ is measurable on S.)

We call (19) the optimality equation for the CTMDP problem (4), and call an $[1,\infty]$-valued lower semianalytic function V a solution to the optimality equation (19) if it satisfies (19) with V^* being replaced by V for each $x\in S$, where $V(x)<\infty$. To guarantee the existence of such a deterministic stationary policy φ as in Theorem 5.1(b), in the next subsection, we shall impose some compactness-continuity conditions, under which the value function will be seen to be measurable or lower semicontinuous.

We postpone the proof of Theorem 5.1 after several lemmas.

Lemma 5.1 *For each $x\in S$ and $\rho\in\mathcal{R}$,*

$$t\in[0,\infty)\to\int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds}V^*(x)$$

is monotone nondecreasing in $t\in[0,\infty)$.

Proof. Let $0 \le t_1 < t_2 < \infty$ be arbitrarily fixed. For the statement of the lemma, it suffices to show

$$\begin{aligned} & \int_0^{t_2} e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^{t_2}(q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) \\ \ge & \int_0^{t_1} e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^{t_1}(q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) \end{aligned} \tag{21}$$

as follows.

Assume that

$$\begin{aligned} \int_0^{t_2} e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau &< \infty; \\ e^{-\int_0^{t_2}(q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) &< \infty. \end{aligned} \tag{22}$$

There is no loss of generality in doing so because otherwise (21) trivially holds.

Then

$$\begin{aligned} & \int_0^{t_2} e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^{t_2}(q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) \\ & - \int_0^{t_1} e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau \\ & - e^{-\int_0^{t_1}(q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) \\ = & \int_{t_1}^{t_2} e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau \\ & + e^{-\int_0^{t_1}(q_x(\rho_s)-c(x,\rho_s))ds} \left(e^{-\int_{t_1}^{t_2}(q_x(\rho_s)-c(x,\rho_s))ds} - 1\right) V^*(x) \\ = & \int_0^{t_2-t_1} e^{-\int_0^{t_1+\tau} (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau \\ & + e^{-\int_0^{t_1}(q_x(\rho_s)-c(x,\rho_s))ds} \left(e^{-\int_{t_1}^{t_2}(q_x(\rho_s)-c(x,\rho_s))ds} - 1\right) V^*(x) \\ = & \int_0^{t_2-t_1} e^{-\int_0^{t_1} (q_x(\rho_s)-c(x,\rho_s))ds} e^{-\int_{t_1}^{t_1+\tau} (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau \\ & + e^{-\int_0^{t_1}(q_x(\rho_s)-c(x,\rho_s))ds} \left(e^{-\int_{t_1}^{t_2}(q_x(\rho_s)-c(x,\rho_s))ds} - 1\right) V^*(x) \\ = & e^{-\int_0^{t_1}(q_x(\rho_s)-c(x,\rho_s))ds} \left\{ \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds} \right. \\ & \left. \times \int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau + e^{-\int_{t_1}^{t_2}(q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) - V^*(x) \right\}. \end{aligned} \tag{23}$$

Let $\delta > 0$ be arbitrarily fixed. By (18), there exists some $\hat{\rho} \in \mathcal{R}$ such that

$$V^*(x) + \delta \ \ge \ \int_0^\infty \int_S V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau) e^{-\int_0^\tau (q_x(\hat{\rho}_s)-c(x,\hat{\rho}_s))ds} d\tau + e^{-\int_0^\infty q_x(\hat{\rho}_s)ds} e^{\int_0^\infty c(x,\hat{\rho}_s)ds}. \tag{24}$$

Define $\tilde{\rho} \in \mathcal{R}$ by

$$\tilde{\rho}_s = \begin{cases} \rho_{t_1+s}, & \text{if } s \le t_2 - t_1; \\ \hat{\rho}_{s-(t_2-t_1)} & \text{if } s > t_2 - t_1. \end{cases} \tag{25}$$

Then

$$\begin{aligned}
V^*(x) &\leq \int_0^\infty e^{-\int_0^\tau (q_x(\tilde{\rho}_s)-c(x,\tilde{\rho}_s))ds}\left(\int_S V^*(y)\tilde{q}(dy|x,\tilde{\rho}_\tau)\right)d\tau + e^{-\int_0^\infty q_x(\tilde{\rho}_s)ds}e^{\int_0^\infty c(x,\tilde{\rho}_s)ds}\\
&= \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau\\
&\quad +\int_{t_2-t_1}^\infty e^{-\int_0^\tau (q_x(\tilde{\rho}_s)-c(x,\tilde{\rho}_s))ds}\left(\int_S V^*(y)\tilde{q}(dy|x,\tilde{\rho}_\tau)\right)d\tau + e^{-\int_0^\infty q_x(\tilde{\rho}_s)ds}e^{\int_0^\infty c(x,\tilde{\rho}_s)ds}\\
&= \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau\\
&\quad +\int_0^\infty e^{-\int_0^{\tau+(t_2-t_1)} (q_x(\tilde{\rho}_s)-c(x,\tilde{\rho}_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\tilde{\rho}_{\tau+t_2-t_1})d\tau + e^{-\int_0^\infty q_x(\tilde{\rho}_s)ds}e^{\int_0^\infty c(x,\tilde{\rho}_s)ds}\\
&= \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau\\
&\quad +\int_0^\infty e^{-\int_0^{t_2-t_1} (q_x(\tilde{\rho}_s)-c(x,\tilde{\rho}_s))ds}e^{-\int_{t_2-t_1}^{\tau+(t_2-t_1)} (q_x(\tilde{\rho}_s)-c(x,\tilde{\rho}_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\tilde{\rho}_{\tau+t_2-t_1})d\tau\\
&\quad +e^{-\int_0^\infty q_x(\tilde{\rho}_s)ds}e^{\int_0^\infty c(x,\tilde{\rho}_s)ds},
\end{aligned}$$

where the first inequality is by (18). Substituting (25) in the last expression, we see

$$\begin{aligned}
V^*(x) &\leq \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau\\
&\quad +e^{-\int_0^{t_2-t_1} (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_0^\infty e^{-\int_{t_2-t_1}^{\tau+(t_2-t_1)} (q_x(\hat{\rho}_{s-(t_2-t_1)})-c(x,\hat{\rho}_{s-(t_2-t_1)}))ds}\\
&\quad \times\int_S V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau + e^{-\int_0^{t_2-t_1} (q_x(\rho_{t_1+s})-c(x,\rho_{t_1+s}))ds}e^{-\int_{t_2-t_1}^\infty q_x(\hat{\rho}_{s-(t_2-t_1)})ds}\\
&\quad \times e^{\int_{t_2-t_1}^\infty c(x,\hat{\rho}_{s-(t_2-t_1)})ds}\\
&= \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau\\
&\quad +e^{-\int_0^{t_2-t_1} (q_x(\rho_{t_1+s})-c(x,\rho_{t_1+s}))ds}\\
&\quad \times\left\{\int_0^\infty e^{-\int_0^\tau (q_x(\hat{\rho}_s)-c(x,\hat{\rho}_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau + e^{-\int_0^\infty q_x(\hat{\rho}_s)ds}e^{\int_0^\infty c(x,\hat{\rho}_s)ds}\right\}\\
&\leq \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau\\
&\quad +e^{-\int_0^{t_2-t_1} (q_x(\rho_{t_1+s})-c(x,\rho_{t_1+s}))ds}V^*(x) + e^{-\int_0^{t_2-t_1} (q_x(\rho_{t_1+s})-c(x,\rho_{t_1+s}))ds}\delta,
\end{aligned}$$

where the last inequality is by (24). Since $\delta > 0$ is arbitrarily fixed, and keeping in mind (22), this amounts to

$$\begin{aligned}
V^*(x) &\leq \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\rho_{s+t_1})-c(x,\rho_{s+t_1}))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_{t_1+\tau})d\tau\\
&\quad +e^{-\int_0^{t_2-t_1} (q_x(\rho_{t_1+s})-c(x,\rho_{t_1+s}))ds}V^*(x).
\end{aligned}$$

This, (22) and (23) imply (21). □

Lemma 5.2 *For each $t \geq 0$ and $x \in S$,*

$$\inf_{\rho\in\mathcal{R}}\left\{\int_0^t e^{-\int_0^s (q_x(\rho_v)-c(x,\rho_v))dv}\int_S V^*(y)\tilde{q}(dy|x,\rho_s)ds + e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds}V^*(x)\right\} = V^*(x).$$

Proof. We only need consider when $t > 0$; the case of $t = 0$ is trivial. Let $\delta > 0$ be arbitrarily fixed. Then by (18), there is some $\hat{\rho} \in \mathcal{R}$ such that

$$V^*(x) + \delta \geq \int_0^\infty e^{-\int_0^\tau (q_x(\hat{\rho}_s)-c(x,\hat{\rho}_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau + e^{-\int_0^\infty q_x(\hat{\rho}_s)ds}e^{-\int_0^\infty c(x,\hat{\rho}_s)ds}.$$

Define $\tilde{\rho} \in \mathcal{R}$ by

$$\tilde{\rho}_s = \hat{\rho}_{t+s},\ \forall\ s > 0.$$

Direct calculations similar to those in the proof of Lemma 5.1 show

$$\begin{aligned} V^*(x)+\delta &\geq \int_0^t e^{-\int_0^\tau (q_x(\hat{\rho}_s)-c(x,\hat{\rho}_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau + e^{-\int_0^t (q_x(\hat{\rho}_s)-c(x,\hat{\rho}_s))ds} \\ &\quad\times\left\{\int_0^\infty e^{-\int_0^\tau (q_x(\tilde{\rho}_s)-c(x,\tilde{\rho}_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\tilde{\rho}_\tau)d\tau + e^{-\int_0^\infty q_x(\tilde{\rho}_s)ds}e^{-\int_0^\infty c(x,\tilde{\rho}_s)ds}\right\} \\ &\geq \int_0^t e^{-\int_0^\tau (q_x(\hat{\rho}_s)-c(x,\hat{\rho}_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau + e^{-\int_0^t (q_x(\hat{\rho}_s)-c(x,\hat{\rho}_s))ds}V^*(x) \end{aligned}$$

where the last inequality is by (18). Since $\delta > 0$ is arbitrarily fixed, the above implies

$$V^*(x) \geq \inf_{\rho\in\mathcal{R}}\left\{\int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds}V^*(x)\right\}.$$

On the other hand, Lemma 5.1 implies

$$V^*(x) \leq \inf_{\rho\in\mathcal{R}}\left\{\int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds}\int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds}V^*(x)\right\}.$$

The statement follows from this and the previous inequality. □

Under extra conditions, including that the function $q_x(a) - c(x,a)$ is bounded, and $0 < \delta < q_x(a) - c(x,a)$ for some constant $\delta > 0$, as in [8], the minimization problem on the right hand side of (18) can be reduced to a problem of Mayer form, and then Lemmas 5.1 and 5.2 follow from Lemma (45.12) of [8].

For the next two lemmas, it is convenient to introduce the following notation. For each $x \in S$ and $T > 0$, let $\mathcal{R}_{V^*,x,T}$ be the set of $\rho \in \mathcal{R}$ such that

$$\int_0^t e^{-\int_0^s (q_x(\rho_v)-c(x,\rho_v))dv}\int_S V^*(y)\tilde{q}(dy|x,\rho_s)ds + e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds}V^*(x) < \infty,\ \forall\ t \in (0,T).$$

Since $V^*(x) \geq 1$, the above inequality is equivalent to

$$\begin{aligned} &\int_0^t e^{-\int_0^s (q_x(\rho_v)-c(x,\rho_v))dv}\int_S V^*(y)\tilde{q}(dy|x,\rho_s)ds < \infty,\ e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds} < \infty,\ \forall\ t \in (0,T); \\ &V^*(x) < \infty, \end{aligned} \tag{26}$$

for each $\rho \in \mathcal{R}_{V^*,x,T}$. Note that if $x \in S$ is such that $V^*(x) < \infty$, then by Lemmas 5.1 and 5.2, $\mathcal{R}_{V^*,x,T} \neq \emptyset$.

Lemma 5.3 *Let $x \in S$ and $T > 0$ be fixed. For each $\rho \in \mathcal{R}_{V^*,x,T}$, it holds that*

$$\int_S V^*(y)\tilde{q}(dy|x,\rho_s) \geq V^*(x)(q_x(\rho_s) - c(x,\rho_s))$$

almost everywhere with respect to $s \in (0,T)$.

Proof. Since $\rho \in \mathcal{R}_{V^*,x,T}$, one can apply the fundamental theorem of calculus and differentiate

$$\int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds} V^*(x)$$

with respect to $t \in (0,T)$, and deduce

$$e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds} \int_S V^*(y)\tilde{q}(dy|x,\rho_t) - e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds}(q_x(\rho_t) - c(x,\rho_t))V^*(x) \geq 0$$

for almost all $t \in (0,T)$ with respect to the Lebesgue measure, where the last inequality is by Lemma 5.1. The statement of the lemma immediately follows. (Recall that (26) holds for each $\rho \in \mathcal{R}_{V^*,x,T}$.) □

Lemma 5.4 *For each $x \in S$, where $V^*(x) < \infty$, (19) is satisfied.*

Proof. Let $T > 0$ be arbitrarily fixed, and so is $x \in S$, where $V^*(x) < \infty$. Then $\mathcal{R}_{V^*,x,T} \neq \emptyset$ as explained earlier. Let some $\rho \in \mathcal{R}_{V^*,x,T}$ be arbitrarily fixed. One can legitimately write

$$\begin{aligned} e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) - V^*(x) &= -\int_0^t (q_x(\rho_\tau) - c(x,\rho_\tau)) e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} d\tau V^*(x), \\ &\quad \forall\ t \in (0,T). \end{aligned}$$

Now,

$$\begin{aligned} &\int_0^t e^{-\int_0^s (q_x(\rho_v)-c(x,\rho_v))dv} \int_S V^*(y)\tilde{q}(dy|x,\rho_s)ds + e^{-\int_0^t (q_x(\rho_s)-c(x,\rho_s))ds} V^*(x) - V^*(x) \\ = &\int_0^t e^{-\int_0^\tau (q_x(\rho_v)-c(x,\rho_v))dv} \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau \\ &-\int_0^t (q_x(\rho_\tau) - c(x,\rho_\tau)) e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} d\tau V^*(x) \\ = &\int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \left\{ \int_S V^*(y)\tilde{q}(dy|x,\rho_\tau) - (q_x(\rho_\tau) - c(x,\rho_\tau))V^*(x) \right\} d\tau \\ = &\int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_A \rho_\tau(da) \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} d\tau \end{aligned}$$

for each $t \in (0,T)$.

By Lemma 5.2, we deduce from the above that

$$\begin{aligned} &0 \\ = &\inf_{\rho \in \mathcal{R}_{V^*,x,T}} \left\{ \int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \int_A \rho_\tau(da) \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} d\tau \right\} \\ \geq &\inf_{\rho \in \mathcal{R}_{V^*,x,T}} \left\{ \int_0^t e^{-\int_0^\tau (q_x(\rho_s)-c(x,\rho_s))ds} \inf_{a \in A} \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} d\tau \right\} \\ \geq &\inf_{\rho \in \mathcal{R}_{V^*,x,T}} \left\{ \int_0^t e^{-\tau \overline{q}_x} \inf_{a \in A} \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} d\tau \right\} \qquad (27) \end{aligned}$$

where the first equality is also because of $V^*(x) < \infty$. Let

$$B(x) = \left\{ a \in A : \int_S V^*(y)\tilde{q}(dy|x,a) < \infty \right\}.$$

Then

$$\begin{aligned} & \inf_{a\in A} \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} \\ = & \inf_{a\in B(x)} \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} \end{aligned} \tag{28}$$

Each $b \in B(x)$ is identified by an element $\rho^b \in \mathcal{R}$ such that $\rho_s^b(da) = \delta_{\{b\}}(da)$ for all $s > 0$. Furthermore, compared with (26) and keeping in mind $V^*(x) < \infty$, we see that $\rho^b \in \mathcal{R}_{V^*,x,T}$. Thus, $\{\rho^b : \; b \in B(x)\} \subseteq \mathcal{R}_{V^*,x,T}$. Hence, for each $a \in B(x)$, one can apply Lemma 5.3, and after that, see

$$\inf_{a\in B(x)} \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} \geq 0.$$

Consequently, we see from (27), (28) and the above inequality that

$$0 \geq \inf_{\rho\in\mathcal{R}_{V^*,x,T}} \left\{ \int_0^t e^{-\tau\overline{q}_x} \inf_{a\in B(x)} \left\{ \int_S V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x) \right\} d\tau \right\} \geq 0,$$

and thus

$$0 = \inf_{a\in A} \left\{ c(x,a)V^*(x) + \int_S q(dy|x,a)V^*(y) \right\},$$

as required. (Recall that $\overline{q}_x < \infty$.) □

Now we are ready to present the proof of Theorem 5.1 as follows.

Proof of Theorem 5.1. Part (a) of this statement has been proved in Lemma 5.4. We prove part (b) of the statement as follows. For each $x \in S$, we can view $\varphi(x)$ as an element of $\mathcal{R}$ by identifying it with $\rho^x \in \mathcal{R}$ such that

$$\rho_t^x(da) := \delta_{\{\varphi(x)\}}(da), \; \forall \; t > 0.$$

Then, $x \in S \to \rho^x$ clearly defines a specific deterministic stationary strategy for the DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ defined in Section 4.

Let $x \in S$ be arbitrarily fixed, where $V^*(x) < \infty$. Then by (20),

$$(q_x(\varphi(x)) - c(x,\varphi(x)))V^*(x) = \int_S \tilde{q}(dy|x,\varphi(x))V^*(y). \tag{29}$$

Note that the right hand side is nonnegative, and is zero if and only if $q_x(\varphi(x)) = 0$, because $V^*(x) \geq 1$ for each $x \in S$. In other words, there are only two possibilities;

$$q_x(\varphi(x)) > c(x,\varphi(x)) \geq 0, \tag{30}$$

or

$$q_x(\varphi(x)) = c(x,\varphi(x)) = 0. \tag{31}$$

In case of (30), we see

$$\begin{aligned} V^*(x) &= \inf_{\rho\in\mathcal{R}}\left\{\int_0^\infty e^{-\int_0^\tau (q_x(\rho_s))-c(x,\rho_s))ds}\left(\int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)\right)d\tau + e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c(x,\rho_s)ds}\right\} \\ &\le \int_0^\infty e^{-\int_0^\tau (q_x(\rho^x_s))-c(x,\rho^x_s))ds}\left(\int_S V^*(y)\tilde{q}(dy|x,\rho^x_\tau)\right)d\tau + e^{-\int_0^\infty q_x(\rho^x_s)ds}e^{\int_0^\infty c(x,\rho^x_s)ds} \\ &= \int_0^\infty e^{-(q_x(\varphi(x))-c(x,\varphi(x)))\tau}(q_x(\varphi(x)) - c(x,\varphi(x)))d\tau V^*(x) + e^{-\int_0^\infty q_x(\varphi(x))ds}e^{\int_0^\infty c(x,\varphi(x))ds} \\ &= V^*(x), \end{aligned}$$

where the first equality is by (18), and the second equality is by (29) and the definition of ρ^x, and the last equality is by (30); recall (1). Thus,

$$\begin{aligned} &\inf_{\rho\in\mathcal{R}}\left\{\int_0^\infty e^{-\int_0^\tau (q_x(\rho_s))-c(x,\rho_s))ds}\left(\int_S V^*(y)\tilde{q}(dy|x,\rho_\tau)\right)d\tau + e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c(x,\rho_s)ds}\right\} \\ &= \int_0^\infty e^{-\int_0^\tau (q_x(\rho^x_s))-c(x,\rho^x_s))ds}\left(\int_S V^*(y)\tilde{q}(dy|x,\rho^x_\tau)\right)d\tau + e^{-\int_0^\infty q_x(\rho^x_s)ds}e^{\int_0^\infty c(x,\rho^x_s)ds} \end{aligned}$$

under (30). Similar calculation show that in case of (31), and in case of $V^*(x)=\infty$, the above equalities hold as well. It remains to apply Proposition 3.3(b); recall the discussions in Section 4 about the reduction of the CTMDP model $\{S,A,q,c\}$ to the DTMDP model $\{\mathbf{X},\mathbf{A},p,l\}$ therein. □

5.2 Existence of a deterministic stationary optimal policy

The objective of this subsection is to show the existence of a deterministic stationary optimal policy for the CTMDP problem (4), under some compactness-continuity conditions.

From now on, the following assumption is always in place.

Assumption 5.1 *For each $x\in S$,*

$$\sup_{a\in A}\{c(x,a)\} =: \overline{c}(x) < \infty.$$

We mention that the function $\overline{c}$ is upper semianalytic on S, and may be not Borel measurable; the similar remark holds for the function $\overline{q}$, see [3]. However, we have the following handy fact, which I learnt from Professor Eugene A. Feinberg.

Lemma 5.5 *There exists a $[1,\infty)$-valued Borel measurable function w on S such that*

$$w(x) \ge 1 + \overline{c}(x) + \overline{q}_x,\ \forall\ x\in S. \tag{32}$$

Proof. This follows from the reasoning in the proof of Lemma 1(a) in [14] based on the Novikov seperation theorem. □

The role of the function w can be also well appreciated in the next subsection.

Next we present two sets of compactness-continuity conditions, under either of which, the main optimality results presented henceforth survive.

Condition 5.1 *(a) The function w from Lemma 5.5 is continuous on S.*

(b) For each bounded continuous function f on S, $\int_S f(y)\tilde{q}(dy|x,a)$ is continuous in $(x,a)\in S\times A$.

(c) The function $c(x,a)$ is lower semicontinuous in $(x,a)\in S\times A$.

(d) *The action space A is a compact Borel space.*

Part (a) of Condition 5.1 is not restrictive, see p.48 of [30].

Condition 5.2 *(a) For each bounded measurable function f on S and each $x \in S$, $\int_S f(y)\tilde{q}(dy|x,a)$ is continuous in $a \in A$.*

(b) *For each $x \in S$, the function $c(x,a)$ is lower semicontinuous in $a \in A$.*

(c) *The action space A is a compact Borel space.*

Condition 5.1 is called the compactness-weak continuity condition, and Condition 5.2 is called the compactness-strong continuity condition. Often, the weak continuity condition is easier for verifications, and it is noted that in some practical applications, the weak continuity condition is satisfied while the strong continuity condition is not, see e.g., Section 6 of [24]. Nevertheless, the two conditions do not imply each other.

Theorem 5.2 *Suppose Condition 5.1 (resp., Condition 5.2) is satisfied. Then there exists a deterministic stationary policy φ satisfying (20) for each $x \in S$, where $V^*(x) < \infty$, and so there exists a deterministic stationary optimal policy φ for the CTMDP problem (4), and the value function V^* is lower semicontinuous (resp., measurable) on S.*

We postpone the proof of Theorem 5.2 after the next few lemmas and preliminaries.

Let us equip $\mathcal{R}$ with the Young topology, which is the weakest topology with respect to which the function

$$\rho \in \mathcal{R} \to \int_0^\infty \int_A f(t,a)\rho_t(da)dt$$

is continuous for each strongly integrable Carathéodory functions f on $(0,\infty) \times A$. Here a real-valued measurable function f on $(0,\infty) \times A$ is called a strongly integrable Carathéodory function if for each fixed $t \in (0,\infty)$, $f(t,a)$ is continuous in $a \in A$, and for each fixed $a \in A$, $\sup_{a\in A} |f(t,a)|$ is integrable in t, i.e., $\int_0^\infty \sup_{a\in A} |f(t,a)|dt < \infty$. See more details in [8].

Lemma 5.6 *Endowed with the Young topology, if the action space A is compact, then $\mathcal{R}$ is a compact Borel space.*

Proof. See Remark 8.2.3 of [1], or Chapter 4 of [8]. □

Lemma 5.7 *(a) Suppose Condition 5.1 is satisfied. Then the DTMDP model $\{\boldsymbol{X}, \boldsymbol{A}, p, l\}$ defined in Section 4 satisfies Condition 3.1.*

(b) *Suppose Condition 5.2 is satisfied. Then the DTMDP model $\{\boldsymbol{X}, \boldsymbol{A}, p, l\}$ defined in Section 4 satisfies Condition 3.2.*

Proof. One can apply the reasoning in the proof of Lemma 3.2 of [29]. □

Lemma 5.8 *Define the stochastic kernel $\tilde{p}$ on $\mathcal{B}(S)$ from $(x,a) \in S \times A$ by*

$$\tilde{p}(dy|x,a) := \frac{q(dy|x,a)}{w(x)} + \delta_{\{x\}}(dy),\ \forall\ (x,a) \in S \times A.$$

Then the following assertions hold.

(a) An $[1,\infty]$-valued lower semianalytic function V on S satisfies

$$0=\inf_{a\in A}\left\{c(x,a)V(x)+\int_S q(dy|x,a)V(y)\right\} \tag{33}$$

for each $x\in S$ such that $V(x)<\infty$ if and only if V is an $[1,\infty]$-valued lower semianalytic solution to

$$V(x)=\inf_{a\in A}\left\{\frac{w(x)}{w(x)-c(x,a)}\int_S \tilde{p}(dy|x,a)V(y)\right\},\ \forall\ x\in S. \tag{34}$$

(b) Let V be an $[1,\infty]$-valued lower semianalytic function on S satisfying (33) for each $x\in S$ such that $V(x)<\infty$. A deterministic stationary policy φ satisfies

$$0=\inf_{a\in A}\left\{c(x,a)V(x)+\int_S q(dy|x,a)V(y)\right\}=c(x,\varphi(x))V(x)+\int_S q(dy|x,\varphi(x))V(y) \tag{35}$$

for each $x\in S$ such that $V(x)<\infty$ if and only if this deterministic stationary policy φ satisfies

$$\inf_{a\in A}\left\{\frac{w(x)}{w(x)-c(x,a)}\int_S \tilde{p}(dy|x,a)V(y)\right\}=\frac{w(x)}{w(x)-c(x,\varphi(x))}\int_S \tilde{p}(dy|x,\varphi(x))V(y),\ \forall\ x\in S.$$

Proof. (a) We first show the "only if" part. Let V be a $[1,\infty]$-valued lower semianalytic solution (33). Let $x\in S$ be fixed. If $V(x)=\infty$, then (34) is satisfied as the both sides are infinite. Suppose now $V(x)<\infty$. Then

$$\begin{aligned}0 &\le c(x,a)V(x)+w(x)\int_S \frac{q(dy|x,a)}{w(x)}V(y)+w(x)V(x)-w(x)V(x)\\ &= c(x,a)V(x)+w(x)\int_S \tilde{p}(dy|x,a)V(y)-w(x)V(x),\ \forall\ a\in A.\end{aligned}$$

Following from this and keeping in mind (32), simple calculations imply

$$V(x)\le\inf_{a\in A}\left\{\frac{w(x)}{w(x)-c(x,a)}\int_S \tilde{p}(dy|x,a)V(y)\right\}.$$

To show the equality, let $\epsilon>0$ be arbitrarily fixed. By (33), there exists some $\hat{a}\in A$ such that

$$\epsilon>c(x,\hat{a})V(x)+\int_S q(dy|x,\hat{a})V(y).$$

The above inequality implies

$$V(x)+\epsilon\ge V(x)+\frac{\epsilon}{w(x)-c(x,\hat{a})}>\frac{w(x)}{w(x)-c(x,\hat{a})}\int_S \tilde{p}(dy|x,\hat{a})V(y),$$

where the first inequality is by (32). Thus, (34) is satisfied.

The similar reasoning applies to show the "if" part; the details are omitted.

(b) This part can be proved as for part (a). □

Lemma 5.9 *(a) Suppose Condition 5.1 is satisfied. Then $(x,a)\in S\times A\to\frac{w(x)}{w(x)-c(x,a)}$ is lower semicontinuous, and for each bounded continuous function f on S, $(x,a)\in S\times A\to\int_S f(y)\tilde{p}(dy|x,a)$ is continuous.*

(b) Suppose Condition 5.2 is satisfied. Then for each $x \in S$, $a \in A \to \frac{w(x)}{w(x)-c(x,a)}$ is lower semicontinuous, and for each bounded measurable function f on S, $a \in A \to \int_S f(y)\tilde{p}(dy|x,a)$ is continuous.

Proof. The statement of this lemma is immediate from Condition 5.1 and Condition 5.2, respectively, as well as the definition of the stochastic kernel $\tilde{p}$. □

Now we are in position to prove Theorem 5.2.

Proof of Theorem 5.2. Suppose Condition 5.1 is satisfied. By Lemma 5.7, one can apply Proposition 3.4(a) to the DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ defined in Section 4. This and Theorem 4.1 imply that the value function V^* of the CTMDP problem (4) is $[1,\infty]$-valued and lower semicontinuous on S. By Theorem 5.1(a), Lemmas 5.8 and 5.9, Proposition 7.31 of [3] and a well known measurable selection theorem, see e.g., Proposition 7.33 of [3], we see that there is a deterministic stationary policy φ for the CTMDP model such that (20) is satisfied for each $x \in S$, where $V^*(x) < \infty$. By Theorem 5.1(b), this deterministic stationary policy φ is optimal for the CTMDP problem (4).

The case when Condition 5.2 is satisfied can be proved in the same way, by applying Proposition 3.4(b) and the corresponding measurable selection theorem, see e.g., Proposition D.5 of [18]. □

5.3 Further reduction to a simpler DTMDP and value iteration

In this subsection, we reduce the CTMDP model $\{S, A, q, c\}$ to a DTMDP $\{S, A, \tilde{p}, \tilde{l}\}$ with the cost function $\tilde{l}$ being defined below. Compared to the DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ defined in Section 4, the DTMDP model here is simpler, with the same state and action space as the original CTMDP model.

Theorem 5.3 *(a) Suppose Condition 5.1 is satisfied. Then the value function V^* for the CTMDP problem (4) is the minimal $[1,\infty]$-valued lower semicontinuous solution to the optimality equation (19).*

(b) Suppose Condition 5.2 is satisfied. Then the value function V^* for the CTMDP problem (4) is the minimal $[1,\infty]$-valued measurable solution to (19).

Proof. (a) Let V be a lower semicontinuous $[1,\infty]$-valued function on S such that (33) is satisfied wherever $V(x) < \infty$. By Proposition 7.33 of [3], there exists a deterministic stationary policy φ satisfying (35) for each $x \in S$ such that $V(x) < \infty$. Let $x \in S$ be fixed. Assume for now $V(x) < \infty$. Arguing as in the proof of Theorem 5.1(b), we see

$$\begin{aligned} & \inf_{\rho \in \mathcal{R}} \left\{ \int_0^\infty e^{-\int_0^\tau (q_x(\rho_s)) - c(x,\rho_s))ds} \left(\int_X V(y)\tilde{q}(dy|x,\rho_\tau) \right) d\tau + e^{-\int_0^\infty q_x(\rho_s)ds} e^{\int_0^\infty c(x,\rho_s)ds} \right\} \\ \leq\ & \int_0^\infty e^{-\tau(q_x(\varphi(x)) - c(x,\varphi(x)))} (c(x,\varphi(x)) - q_x(\varphi(x))) d\tau V(x) + e^{-\int_0^\infty q_x(\varphi(x))ds} e^{\int_0^\infty c(x,\varphi(x))ds} \\ \leq\ & V(x), \end{aligned}$$

where for the last inequality, recall that one only needs deal with two possibilities, namely, (30) and (31), because of (35). The previous inequality holds trivially if $V(x) = \infty$. Now by applying Proposition 3.3 to the DTMDP model $\{\mathbf{X}, \mathbf{A}, p, l\}$ in Section 4, c.f., (18), as well as Theorem 4.1, we see $V(x) \geq V^*(x)$ for each $x \in S$. This and Theorem 5.1(a) imply part (a) of this statement.

(b) This part can be proved in the same way as for part (a) by using the appropriate measurable selection theorem, c.f. Proposition D.5 of [18]. □

Now we are in position to present the equivalent DTMDP model $\{S, A, \tilde{p}, \tilde{l}\}$.

Define for each $(x, a, y) \in S \times A \times S$,

$$\tilde{l}(x, a, y) := \ln \frac{w(x)}{w(x) - c(x, a)}.$$

Recall by (32), $\tilde{l}(x, a, y) > 0$ for each $(x, a, y) \in S \times A \times S$.

Consider the DTMDP model $\{S, A, \tilde{p}, \tilde{l}\}$ (with the exponential utility). Note that (34) is the optimality equation for this DTMDP model. Suppose Condition 5.1 or Condition 5.2 is satisfied. Then by Theorem 5.3, Proposition 3.4 applied to the DTMDP model $\{S, A, \tilde{p}, \tilde{l}\}$, and Lemma 5.8(b), we see the DTMDP model $\{S, A, \tilde{p}, \tilde{l}\}$ and the CTMDP model $\{S, A, q, c\}$ are equivalent; the value functions are the same, and a deterministic stationary optimal policy for the CTMDP model gives a deterministic stationary optimal strategy for the DTMDP model $\{S, A, \tilde{p}, \tilde{l}\}$, and vice versa.

As a consequence, we can write down the value iteration algorithm for the CTMDP problem (4).

Corollary 5.1 *Suppose Condition 5.1 or Condition 5.2 is satisfied. Define $V^{(0)}(x) := 1$ for each $x \in S$, and for each $n = 1, 2, \ldots,$*

$$V^{(n)}(x) = \inf_{a \in A} \left\{ \int_S \tilde{p}(dy|x, a) e^{\tilde{l}(x,a,y)} V^{(n-1)}(y) \right\}, \ \forall\ x \in S.$$

Then for each $x \in S$, $V^{(n)}(x)$ increases to $V^*(x)$ as $n \uparrow \infty$, where V^* is the value function of the CTMDP problem (4).

Proof. The statement follows from the discussions above this theorem, and Proposition 3.4. Recall Lemma 5.9. □

6 Conclusion

To sum up, for the CTMDP problem, where the certainty equivalent with respect to the exponential utility of the total undiscounted cost is to be minimized, we established the optimality equation. Under the compactness-continuity condition, we showed the existence of a deterministic stationary optimal policy. By investigating the optimality equation, we reduced the CTMDP problem to an equivalent DTMDP problem, which is with the same state and action space as the original CTMDP. In particular, the value iteration algorithm for the CTMDP problem follows from this reduction. Note that, we did not need impose any condition on the growth of the transition rate, and the cost rate is unbounded, and the controlled process in the Borel state space could be explosive.

As for applications, we believe that our results will be useful for optimal control of queueing systems. In fact, [6] considered the risk-sensitive control of a queueing system as a CTMDP with the total discounted cost in a finite state and action space. Restricted to deterministic stationary policies, although the authors of [6] applied the uniformization technique to reduce the CTMDP to a DTMDP, they did not manage to investigate the continuous-time problem because the resulting DTMDP was nonstandard. The reduction method here is different from the uniformization technique, and its application to the discounted problem will be more delicate.

Acknowledgement. This work was carried out with a financial grant from the Research Fund for Coal and Steel of the European Commission, within the INDUSE-2-SAFETY project (Grant No. RFSR-CT-2014-00025). I would like to thank Alexey B. Piunovskiy for the helpful discussions. Finally, I would like to thank the referees for the careful reading and useful remarks.

References

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

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

[3] Bertsekas, D. and Shreve, S. (1978). *Stochastic Optimal Control.* Academic Press, New York.

[4] Cavazos-Cadena, R. and Montes-de-Oca, R. (2000). Optimal stationary policies in risk-sensitive dynamic programs with finite state space and nonnegative rewards. *Appl. Math. (Warsaw)* **27**, 167-185.

[5] Chung, K. and Sobel, M. (1987). Discounted MDP's: distribution functions and exponential utility maximization. *SIAM J Control Optim.* **25**, 49-62.

[6] Coraluppi, S. and Marcus, S. (1997). Risk-sensitive queueing. *Proceedings of the 35th Annual Allerton Conference on Communication Control and Computing*, 943-952.

[7] Costa, O. and Dufour, F. (2013). *Continuous Average Control of Piecewise Deterministic Markov Processes.* Springer, New York.

[8] Davis, M. (1993). *Markov Models and Optimization.* Chapman and Hall, London.

[9] Dynkin, E. and Yushkevich, A. (1979). *Controlled Markov Processes.* Springer, New York.

[10] Fainberg, E. (1982). Controlled Markov processes with arbitrary numerical criteria. *Theory Probab. Appl.* **27**, 486-503.

[11] Feinberg, E. (1996). On measurability of value function and representation of randomized policies in Markov decision processes. In *Statistics, Probability and Game Theory Papers in Honor of David Blackwell*, Ferguson, T. et al.(eds): 29-43, Institute of Mathematical Statistics, Hayward.

[12] Feinberg, E. (2004). Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Oper. Res.* **29**, 492-524.

[13] Feinberg, E. (2005). On essential information in sequential decision processes. *Math. Meth. Oper. Res.* **62**, 399-410.

[14] Feinberg, E., Mandava, M. and Shiryaev, A. (2016) Kolmogorov's equations for jump Markov processes with unbounded jump rates. Preprint. Available at arXiv:1603.02367.

[15] Forwick, L., Schäl, M. and Schmitz, M. (2004). Piecewise deterministic Markov control processes with feedback controls and unbounded costs. *Acta Appl. Math.* **82**, 239-267.

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

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

[18] Hernández-Lerma, O. and Lasserre, J. (1996). *Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[19] Howard, R. and Matheson, J. (1972). Risk-sensitive Markov decision proceses. *Manag. Sci.* **18**, 356-369.

[20] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nykodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebite.* **31**, 235-253.

[21] Jaquette, S. (1976). A utility criterion for Markov decision processes. *Manag. Sci.* **23**, 43-49.

[22] Jaśkiewicz, A. (2008). Average optimality for risk-sensitive control with general state space. *Ann. Appl. Probab.* **17**, 654-675.

[23] Jaśkiewicz, A. (2008). A note on negative dynamic programming for risk-sensitive control. *Oper. Res. Lett.* **36**, 531-534.

[24] Jaśkiewicz, A. (2009). Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities. *J. Optim. Theory Appl.* **141**, 321-347.

[25] Kitaev, M. (1986). Semi-Markov and jump Markov controlled models: average cost criterion. *Theory. Probab. Appl.* **30**: 272-288.

[26] Kitaev, M. and Rykov, V. (1995). *Controlled Queueing Systems.* CRC Press, Boca Raton.

[27] Piunovski, A. and Khametov, V. (1985). New effective solutions of optimality equations for the controlled Markov chains with continuous parameter (the unbounded price-function). *Problems Control Inform. Theory* **14**, 303-318.

[28] Piunovskiy, A. (1997). *Optimal Control of Random Sequences in Problems with Constraints*, Kluwer, Dordrecht.

[29] Schäl, M. (1998). On piecewise deterministic Markov control processes: control of jumps and of risk processes in insurance. *Insur. Math. Econ.* **22**, 75-91.

[30] Srivastava, S. (1998). *A Course on Borel Sets.* Springer, New York.

[31] Wagner, W. (2005). Explosion phenomena in stochastic coagulation-fragmentation models. *Ann. Appl. Probab.* **15**, 2081-2112.

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

[33] Yushkevich, A. (1980). On reducing a jump controllable Markov model to a model with discrete time. *Theory. Probab. Appl.* **25**, 58-68.