UNIVERSITY OF LIVERPOOL

Risk-sensitive Markov Decision Processes

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy.

Xin Guo

201175934

Supervised by Dr. Yi Zhang, Dr. Alexey Piunovskiy

Department of Mathematical Sciences
University of Liverpool

Signature _____

Date _____ / _____ / _____

April 12, 2020

Statement of Originality

This dissertation was written by myself, in my own words, except for quotations from published and unpublished sources which are clearly indicated. I am conscious that the incorporation of material from other works or a paraphrase of such material without acknowledgement will be treated as plagiarism, according to the University Academic Integrity Policy. The source of any picture, map or other illustration is also indicated, as is the source, published or unpublished, of any material not resulting from my own research.

Acknowledgements

My sincere gratitude goes first and foremost to my supervisor, Dr. Yi Zhang, for his constant encouragement and instructive guidance on my thesis. He has provided me with patient instruction through all the stages of writing the thesis. I am deeply grateful of his help in the completion of this thesis. Besides, I would also like to thank Dr. Alexey Piunovskiy, my secondary supervisor, for his detailed materials helping me strengthen my basic knowledge of the thesis.

I would like to express my heartfelt gratitude to all the teachers and staffs at the Department of Mathematical Sciences, who provide me with a comfortable place to study and research. I also owe my special gratitude to my friends and colleagues for their kind help and companion in the past four years.

Special thanks to my beloved family for their continuous support and great confidence in me all through these years.

Abstract

In this thesis, we mainly consider continuous-time Markov decision processes (CTMDPs) with risk-sensitive case and other applications. In an effort to extend the cost/reward rates to be unbounded, we may weaken some conditions compared with previous articles, or use another method to make the main results better in some way. Chapter 1 is a general introduction to continuous-time Markov decision problems. For risk-sensitive problems, chapter 2 to chapter 5 make a detailed discussion. where finite horizon case, average case, gradual-impulse case and piecewise case are included. The last two chapters are about other problems in CTMDPs.

Chapter 2 considers a risk-sensitive CTMDP over a finite time duration. Under the conditions that can be satisfied by unbounded transition and cost rates, we show the existence of an optimal policy, and the existence and uniqueness of the solution to the optimality equation out of a class of possibly unbounded functions, to which the modified Feynman-Kac formula was also justified to hold.

Chapter 3 is about risk-sensitive average optimization for denumerable CTMDPs, in which the transition and cost rates are allowed to be unbounded, and the policies can be randomized history-dependent. Based on the results obtained in last chapter and some new properties, we establish the existence and uniqueness of a solution to the risk-sensitive average optimality equation (RS-AOE), and also prove the existence of an optimal stationary policy via the RS-AOE and the extended Feymanm-Kac's formula. Furthermore, for the case of finite actions available at each state, we construct a sequence of models of finite-state CTMDPs with optimal stationary policies which can be obtained by a policy iteration algorithm in a finite number of iterations, and prove that an average optimal policy for the case of infinitely countable states can be approximated by those of the finite-state models.

In chapter 4, the risk-sensitive gradual-impulse control problem of CTMDPs

is studied. We prove, under very general conditions on the system primitives, the existence of a deterministic stationary optimal policy out of a more general class of policies. Policies that we consider allow multiple simultaneous impulses, randomized selection of impulses with random effects, relaxed gradual controls, and accumulation of jumps. After characterizing the value function using the optimality equation, we reduce the continuous-time gradual-impulse control problem to an equivalent simple discrete-time Markov decision process, whose action space is the union of the sets of gradual and impulsive actions.

Chapter 5 discusses piecewise deterministic Markov decision process (PDMDP), where the expected exponential utility of total (nonnegative) cost is to be minimized. The cost rate, transition rate and post-jump distributions are under control. Under natural conditions, we establish the optimality equation, justify the value iteration algorithm, and show the existence of a deterministic stationary optimal policy. Applied to special cases, the obtained results already significantly improve some existing results in the literature on finite horizon and infinite horizon discounted risk-sensitive CTMDPs.

After risk-sensitive problems, Chapter 6 talks about discounted CTMDPs, where the negative part of each cost rate is bounded by a drift function, say w, whereas the positive part is allowed to be arbitrarily unbounded. Our focus is on the existence of a stationary optimal policy for the discounted CTMDP problems out of the more general class. Both constrained and unconstrained problems are considered. As a consequence, we withdraw and weaken several conditions commonly imposed in the literature.

And the last chapter 7 is an application of CTMDPs, a two-person zero-sum continuous-time Markov pure jump game in Borel state and action spaces over a fixed finite horizon. The main assumption on the model is the existence of a drift function, which bounds the reward rate. Under some regularity conditions, we show that the game has a value, and both of the players have their optimal policies.

Contents

Notation

a	A generic action for gradual control / for maximizer
b	A generic action for impulse control / for minimizer
i	A generic state for a denumerable state space model
x	A generic state for a Borel state space model
$\mathcal{B}(\mathbf{X})$	The Borel σ-algebra on a Borel space $\mathbf{X}$
$B_V(\mathbf{X})$	The space of all V-bounded functions on $\mathbf{X}$
$\mathbb{C}_b(\mathbf{X})$	The space of all bounded continuous functions on $\mathbf{X}$
C^1_{V,V_1}	The space of all continuous V-bounded φ on $[0,T]$ with V_1-bounded φ'
$\mathbb{E}^\sigma_x$	Expectation wrt the strategic measure of the DTMDP under the strategy σ
J	The risk-sensitive average CTMDP criterion
$\mathbb{K}$	Set of all feasible state-action pairs
$\mathcal{L}$	The risk-sensitive gradual-impulse CTMDP criterion
$\mathbb{P}(\mathbf{X})$	Space of probability measures on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ endowed with weak topology
$\mathcal{R}$	Collection of $\mathbb{P}(\mathbf{X})$-valued measurable mappings
$\mathcal{V}$	The finite horizon risk-sensitive CTMDP criterion

1 Continuous-time Markov decision processes

In this chapter we formally introduce the precise definitions of state and action processes in continuous-time Markov decision processes (CTMDP), some fundamental properties, and the basic optimality criteria that we are interested in.

1.1 Introduction

Notation: Given a Borel space $\mathbf{X}$, its Borel σ-algebra is denoted by $\mathcal{B}(\mathbf{X})$. By convention, when referring to sets or functions, "measurable" means "Borel-measurable." we denote by $\mathbb{C}_b(\mathbf{X})$ the space of all bounded continuous functions on $\mathbf{X}$

Given any $T > 0$, for each measurable function ψ on $[0, T] \times \mathbf{X}$, if $\psi(\cdot, x)$ is absolutely continuous on $[0, T]$, then we put ψ' a measurable function on $[0, T] \times \mathbf{X}$ such that $\psi(t, x) - \psi(0, x) = \int_0^t \psi'(s, x)ds$ for each $x \in \mathbf{X}$ and $t \in [0, T]$.

For any measurable function $V, V_1 \geq 1$ on $\mathbf{X}$, we define the V-weighted supremum norm $\|\cdot\|_V$ of a real-valued measurable function φ on $[0, T] \times \mathbf{X}$ by

$$\|\varphi\|_V := \sup_{(t,x)\in[0,T]\times\mathbf{X}} \{\frac{|\varphi(t, x)|}{V(x)}\},$$

we call the function φ V-bounded if the norm is finite, and $C^1_{V,V_1}([0, T] \times \mathbf{X})$ is the collection of V-bounded functions $\varphi(t, x)$ on $[0, T] \times \mathbf{X}$ such that $\varphi(t, x)$ is absolutely continuous on $[0, T]$ for each x in $\mathbf{X}$, which admits some V_1-bounded φ', and define $B_V(\mathbf{X}) := \{\varphi : \quad \|\varphi\|_V < \infty\}$.

We adopt the conventions of

$$\frac{0}{0} := 0,\ 0 \cdot \infty := 0,\ \frac{1}{0} := +\infty,\ \infty - \infty := \infty. \tag{1.1}$$

1.2 The general control model

The control model associated with the CTMDP that we are concerned with is a five-tuple

$$\mathcal{M} = \{\mathbf{S}, \mathbf{A}, \mathbf{A}(\cdot,\cdot), q(dy|t,x,a), c(t,x,a)\} \tag{1.2}$$

with the following components:

(a) a Borel set $\mathbf{S}$, called the *state space*, which is the set of all states of the system under observation;

(b) a Borel space $\mathbf{A}$, called the *action space*;

(c) a family $(\mathbf{A}(t,x), t \in [0,\infty), x \in \mathbf{S})$ of nonempty measurable subsets $\mathbf{A}(t,x)$ of $\mathbf{A}$, where $\mathbf{A}(t,x)$ denotes the set of actions or decisions available to the controller when the state of the system is $x \in \mathbf{S}$. Let

$$\mathbb{K} := \{(t,x,a) | x \in \mathbf{S}, a \in \mathbf{A}(t,x)\} \tag{1.3}$$

be the set of all feasible state-action pairs.

(d) the transition rates $q(dy|t,x,a)$ is a signed kernel defined on $\mathcal{B}(\mathbf{S})$ given $(t,x,a) \in \mathbb{K}$ such that $\widetilde{q}(\Gamma|t,x,a) := q(\Gamma \setminus \{x\}|t,x,a) \geq 0$ for all $\Gamma \in \mathcal{B}(\mathbf{S})$. Throughout this thesis, we assume that $q(\cdot|t,x,a)$ is conservative and stable, i.e.,

$$q(\mathbf{S}|t,x,a) = 0, \ \bar{q}_x = \sup_{a \in \mathbf{A}(t,x)} q_x(a) < \infty, \tag{1.4}$$

where $q_x(a) := -q(\{x\}|t,x,a)$.

(e) a measurable real-valued function $c(t,x,a)$ on $\mathbb{K}$, called the *cost function*, which is assumed to be measurable in $a \in \mathbf{A}(t,x)$ for each fixed $t \geq 0$ and

$x \in \mathbf{S}$. (As $c(t,x,a)$ is allowed to take positive and negative values, it can also be interpreted as *reward function*).

Now we describe the construction of CTMDP. Let us take the sample space Ω by adjoining to the countable product space $\mathbf{S}\times((0,\infty)\times\mathbf{S})^\infty$ the sequences of the form $(x_0,\theta_1,\dots,\theta_n,x_n,\infty,x_\infty,\infty,x_\infty,\dots)$, for some $n\geq 0$, where $x_0,x_1,\dots,x_n$ belong to $\mathbf{S}$, $\theta_1,\dots,\theta_n$ belong to $(0,\infty)$, and $x_\infty\notin\mathbf{S}$ is the isolated point. Below we denote $\mathbf{S}_\infty := \mathbf{S}\cup\{x_\infty\}$ We equip Ω with its Borel σ-algebra $\mathcal{F}$.

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

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

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 [71] for greater details. The marked point process $\{t_n,x_n\}$ defines the stochastic process on $(\Omega,\mathcal{F})$ of interest $\{\xi_t,t\geq 0\}$ by

$$\xi_t(\omega)=\sum_{n\geq 0} I\{t_n\leq t<t_{n+1}\}x_n + I\{t_\infty\leq t\}x_\infty. \tag{1.5}$$

Here we accept $0\cdot x:=0$ and $1\cdot x:=x$ for each $x\in\mathbf{S}_\infty$.

Definition 1.1. *(a) A (history-dependent) policy π is determined and often identified by a sequence of stochastic kernels $\{\pi_n, n=0,1,\dots\}$ such that*

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

where $a_\infty\notin \boldsymbol{A}$ is some isolated point. For each n, $\pi_n(da|x_0,\theta_1,\dots,x_n,s)$ is a s-

tochastic kernel concentrated on $\boldsymbol{A}(t_n+s, x_n)$ *given* $x_0 \in \boldsymbol{S}$, $\theta_1 \in (0,\infty), \ldots,$ $x_n \in \boldsymbol{S}$, $s \in (0,\infty)$. *We identify a policy* π *with the sequence of stochastic kernels* $\{\pi_n\}_{n=0}^{\infty}$.

(b) A policy π *is called Markov if, for some stochastic kernel* π^M *on* $\boldsymbol{A}$ *concentrated on* $\boldsymbol{A}(t,x)$ *from* $(x,t) \in \boldsymbol{S} \times (0,\infty)$, *one can write* $\pi(da|\omega,t) = \pi^M(da|\xi_{t-},t)$ *whenever* $t < t_\infty$. *A Markov policy is identified with the underlying stochastic kernel* π^M. *A Markov policy* π^M *is called deterministic if there exists a measurable function* $f(t,i)$ *on* $[0,\infty] \times \boldsymbol{S}$ *such that* $\pi^M(da|i,t) = \delta_{\{f(t,i)\}}(da)$

(c) A policy $\pi = \{\pi_n\}_{n=0}^{\infty}$ *is called stationary if, with slight abuse of notations,*

$$\pi_n(da|x_0,\theta_1,\ldots,x_n,s) = \pi(da|x_n)$$

for each of the stochastic kernels π_n. *A stationary policy is further called deterministic if* $\pi(da|x) = \delta_{\{f(x)\}}(da)$ *for some measurable mapping* f *from* $\boldsymbol{S}$ *to* $\boldsymbol{A}$ *such that* $f(x) \in \boldsymbol{A}(t,x)$ *for each* $x \in \boldsymbol{S}$. *We shall identify such a deterministic stationary policy with the underlying measurable mapping* f.

The class of all policies for the CTMDP is denoted by Π, and the class of all Markov policies is Π_m^r. We also denote by Π_m^d the set of deterministic Markov policies, by F the set of all stationary policies.

For each $\pi \in \Pi$, the random measure m^π defined by

$$m^\pi(j|\omega,t)dt := \int_A q(j \setminus \{\xi_{t-}\}|t,\xi_{t-},a)\pi(da|\omega,t)dt \tag{1.6}$$

is predictable, see [64].

For any initial distribution γ on $\mathbf{S}$ and policy $\pi \in \Pi$, the Ionescu Tulcea theorem ensures the existence of a unique probability measure P_γ^π on $(\Omega,\mathcal{F})$ in [54, 55]. The following facts show how the initial distribution and transition probabilities can decide the probability P_x^π on $(\Omega,\mathcal{F})$: for any $C_n \in \mathcal{B}(A)$ and $E_n \in \mathcal{B}(S)$, as well as $n \geq 0$, we have

(1) $P_x^\pi(x_0 = x) = 1$;

(2) $P_x^\pi(x_n \in E_n | x_0, a_0, \ldots, x_{n-1}, a_{n-1}) = q(E_n | x_{n-1}, a_{n-1})$ for $n \geq 1$;

(3) $P_x^\pi(a_n \in C_n | h_n) = \pi_n(C_n | h_n)$;

(4) $P_x^\pi(a_0 \in C_0, \ldots, x_n \in E_n, a_n \in C_n, x_{n+1} \in E_{n+1}) = \int_{C_0} \pi_0(da_0|x)$
$\int_{E_1} q(dx_1|x_0, a_0) \cdots \int_{C_n} \pi_n^1(da_n|h_n) q(E_{n+1}|x_n, a_n)$.

Let E_γ^π be its corresponding expectation operator. In particular, E_γ^π and P_γ^π will be respectively written as E_x^π and P_x^π when γ is the Dirac measure located at a state x in $\mathbf{S}$.

Then we introduce some further notations. $\mathbb{P}(\mathbf{A})$ stands for the space of probability measures on $(\mathbf{A}, \mathcal{B}(\mathbf{A}))$. We endow $\mathbb{P}(\mathbf{A})$ with its weak topology (generated by bounded continuous functions on $\mathbf{A}$) and the Borel σ-algebra, so that $\mathbb{P}(\mathbf{A})$ is a Borel space, see Chapter 7 of [9]. Let $\mathcal{R}$ be the collection of $\mathbb{P}(\mathbf{A})$-valued measurable mappings on $[0, \infty)$ with any two elements therein being identified the same if they differ only on a null set with respect to the Lebesgue measure. It is known, see Lemma 1 of [104], that the space $\mathcal{R}$, endowed with the smallest σ-algebra with respect to which the mapping $\rho = (\rho_t(da)) \in \mathcal{R} \to \int_0^\infty e^{-t} g(t, \rho_t) dt$ is measurable for each bounded measurable function g on $(0, \infty) \times \mathbb{P}(\mathbf{A})$, is a Borel space. Then, according to Section 43 of [23], the space $\mathcal{R}$ is a compact metrizable space, endowed with the Young topology when A is compact, which is the coarsest topology with respect to which, the mapping

$$\rho = (\rho_t(da)) \in \mathcal{R} \to \int_0^\infty \int_{\mathbf{A}} g(t, a) \rho_t(da) dt$$

is continuous for each function g on $(0, \infty) \times \mathbf{A}$ satisfying that (a) for each $t \in (0, \infty)$, $g(t, \cdot)$ is continuous on $\mathbf{A}$; (b) for each $a \in \mathbf{A}$, $g(\cdot, a)$ is measurable on $(0, \infty)$; and (c) $\int_0^\infty \sup_{a \in \mathbf{A}} |g(t, a)| dt < \infty$. Such a function g satisfying these requirements is called a strongly integrable Caratheodory function.

For each $\mu \in \mathbb{P}(\mathbf{A})$, we denote

$$q_x(\mu) := \int_{\mathbf{A}} q_x(a)\mu(da),\ \tilde{q}(dy|x,\mu) := \int_{\mathbf{A}} \tilde{q}(dy|x,a)\mu(da)$$
$$c(x,\mu) := \int_{\mathbf{A}} c(x,a)\mu(da).$$

1.3 Risk-sensitive problems

Before stating the optimality criteria, we spend some time talking about the risk-sensitive problems, which we have to deal with in most of our thesis. We consider the performance of CTMDP measured by the expectation of the exponential utility of the total cost. Such problems are often called risk-sensitive (RS) because they take both the first order (expectation) and higher order moments. We give the following formal justifications as in [12, 43].

Let X be the (possibly random) reward, and so the concerned performance measure is $E[e^{\theta X}]$, where $\theta > 0$ is a fixed constant. Let $ceq(X) := \frac{1}{\theta}\ln(E[e^{\theta X}])$ be a (deterministic) constant such that $E[e^{\theta X}] = e^{\theta ceq(X)}$. (For simplicity, assume all the involved expectations are finite.) Then applying the Taylor expansion around $E[X]$:

$$e^{\theta X} \approx e^{\theta E[X]} + \theta e^{\theta E[X]}(X - E[X]) + \frac{1}{2}\theta^2 e^{\theta E[X]}(X - E[X])^2$$
$$e^{\theta ceq(X)} \approx e^{\theta E[X]} + \theta e^{\theta E[X]}(ceq(X) - E[X]);$$

(where the second function uses the fact of Taylor expansion of $\ln(1+x)$ to get $\ln(1+x) \approx x$ and we take $x := \theta(ceq(x) - E(x)))$, so that $e^{\theta ceq(X)} = E[e^{\theta X}] \approx e^{\theta E[X]} + \frac{1}{2}\theta^2 e^{\theta E[X]} Var(X)$ by taking expectation on the both sides of the second equality in the above. Comparing this with the first equality in the above, we see $ceq(X) - E[X] \approx \frac{1}{2}\theta Var(X) \geq 0$.

Thus, the performance of $E[e^{\theta X}]$ takes into account both $E[X]$ and $Var(X)$

compared to the case of linear utility where only $E[X]$ is counted. We also mention that $ceq(X)$ is the so called certainty equivalent of X, which can be interpreted as the reward the controller decides to accept. Thus $ceq(X) - E[X] \geq 0$ means the decision maker is risk-averse, that's to say the decision maker will accept the certain reward only when it is more than expected. Below, we will regard $\theta = 1$ without loss of generality. The CTMDP with a linear utility is called risk-neutral. Risk-sensitive and risk-neutral problems might admit quite different optimality results in general. For example, in a model with finite state and action space, there is always an optimal deterministic stationary policy for discounted risk-neutral CTMDPs, whereas this is not the case for the risk sensitive counterpart, see [43].

1.4 Optimality criteria

After the preliminaries and introduction of risk-sensitive problems above, we now define several optimality criteria, some of which do not consider the change of time t and are called *homogeneous* while the others are called *non-homogeneous* models. When referring to homogeneous models we just omit t in the previous notations like $\mathbf{A}(x), q(dy|x,a), c(x,a)$ etc. Criteria listed below are what we are interested, the risk-sensitive finite horizon CTMDP, risk-sensitive average CTMDP problems, risk-sensitive gradual-impulse CTMDP, risk-sensitive piecewise deterministic Markov decision processes (PDMDP) and the expected discounted CTMDP problem.

We have to note that in the following chapters, when the state space $\mathbf{S}$ is denumerable, we use $i, j, \cdots$ to denote the states for convenience.

Definition 1.2. (The finite horizon nonhomogeneous RS-CTMDP criterion)

$$\mathcal{V}(\pi, i) := E_i^{\pi}\left[e^{\int_0^T \int_{\mathbf{A}} c(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_T)}\right]. \tag{1.7}$$

defines the performance measure. For each $x \in \mathbf{S}$, let

$$\mathcal{V}^*(i) = \inf_{\pi \in \Pi} \mathcal{V}(\pi, i) = \mathcal{V}(\pi^*, i).$$

where the policy $\pi^* \in \Pi$ is said to be optimal.

Definition 1.3. (The risk-sensitive average CTMDP criterion)

$$J(i, \pi) := \limsup_{T \to \infty} \frac{1}{T} \ln E_i^{\pi} \left[e^{\int_0^T \int_{\mathbf{A}} c(\xi_t, a) \pi(da|\omega, t) dt} \right] \tag{1.8}$$

for each $i \in \mathbf{S}$ and $\pi \in \Pi$.

A policy $\pi^* \in \Pi$ is said (risk-sensitive average) optimal if for all $i \in \mathbf{S}$

$$J(i, \pi^*) = \inf_{\pi \in \Pi} J(i, \pi)$$

Definition 1.4. (The risk-sensitive gradual-impulse CTMDP criterion)

$$\mathcal{L}(u, x) \;\; := \;\; E_x^u \left[e^{\sum_{n=1}^{\infty} \left(C^I(Y_n) + \int_{T_n}^{T_{n+1}} \int_{\mathbf{A}^G} c^G(\bar{x}(\xi_s), a) \Pi_n(da|H_n, s - T_n) ds \right)} \right]$$

A policy u^* satisfying $\mathcal{L}(x, u^*) = \mathcal{L}^*(x)$ for all $x \in \mathbf{S}$ is called optimal for the gradual-impulse control problem:

$$\text{Minimize over } u \in \mathcal{U} : \mathcal{L}(x, u). \tag{1.9}$$

The exact meaning of notations in the gradual-impulse model can be referred in Chapter 5.

Definition 1.5. (The risk-sensitive PDMDP criterion)

It is assumed that for each $x \in \mathbf{S}$

$$\phi(x, t + s) = \phi(\phi(x, t), s), \;\; \forall \; s, t \geq 0; \;\; \phi(x, 0) = x, \tag{1.10}$$

For each $x \in \mathbf{S}$, and policy $\pi = (\pi_n)$,

$$V(x,\pi) = E_x^{\pi}\left[e^{\sum_{n=0}^{\infty}\int_0^{\theta_{n+1}}\int_{\mathbf{A}} c(\phi(x_n,s),a)\pi_n(da|x_0,\theta_1,\ldots,x_n,s)ds}\right]$$

A policy π^* is called optimal if $V(x,\pi^*) = \inf_{\pi\in\Pi} V(x,\pi) =: V^*(x)$.

Definition 1.6. (The expected discounted CTMDP criterion)

$$W_\alpha(x,\pi) = E_x^{\pi}\left[\int_0^{\infty} e^{-\alpha t}\int_{\mathbf{A}} c(\xi_t,a)\pi(da|\omega,t)dt\right], \tag{1.11}$$

defines the concerned performance measure of the policy $\pi \in \Pi$ given the initial state $x \in \mathbf{S}$ and fixed discount factor $\infty > \alpha > 0$.

The corresponding optimal value function of the problem is

$$W_\alpha^*(x) := \inf_{\pi\in\Pi} W_\alpha(x,\pi) = W_\alpha^*(x,\pi^*)$$

Part I

Risk-sensitive problems

2 Finite horizon risk-sensitive CTMDP with unbounded rates

2.1 Introduction

In this chapter, we consider a risk-sensitive continuous-time Markov decision process over a finite time duration. From the results of chapter 5 about the PDMDP, it is naturally to think that whether we can extend the finite horizon CTMDP problem with nonnegative cost rates to the unbounded case. At the same time, considering discounted CTMDP problem with a lower bounding function in chapter 6, where the technique used there is a transformation from general case to the nonnegative cost rate. If we can use the similar transformation, then it is just an application of the risk-sensitive PDMDP results. Unfortunately, we still don't know how to combine these two ways together to get what we want for the finite horizon risk-sensitive CTMDP with unbounded cost rates, so we change a way to look for the modified Feyman-Kac formula to get the results. In the following, under the conditions that can be satisfied by unbounded transition and cost rates, we show the existence of an optimal policy, and the existence and uniqueness of the solution to the optimality equation out of a class of possibly unbounded functions, to which the Feynman-Kac formula was also justified to hold.

2.2 Conditions and statements

In this section, we impose a set of conditions allowing one to consider unbounded transition and cost rates, see Example 2.1 below, and present several preliminary

statements, which will serve the proof of Theorem 2.2 below.

First we recall the definition of the corresponding criteria and give some conditions that should be satisfied:

$$\mathcal{V}(\pi, i) := E_i^\pi \left[e^{\int_0^T \int_{\mathbf{A}} c(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_T)} \right].$$

Condition 2.1. *There exist a $[1,\infty)$-valued function V defined on $\boldsymbol{S}$ and constants $\rho > 0$, $M > 1$ such that*

(a) $\sum_{j \in \boldsymbol{S}} q(j|t,i,a)V(j) \le \rho V(i)$ *for each* $(t,i,a) \in \mathbb{K}$;

(b) $\bar{q}_i \le MV(i)$ *for all* $i \in \boldsymbol{S}$;

(c) $e^{2(1+T)|c(t,i,a)|} \le MV(i)$ *for each* $(t,i,a) \in \mathbb{K}$, *and* $e^{2(1+T)|g(i)|} \le MV(i)$ *for each* $i \in \boldsymbol{S}$. *(For the case of* $g(i) \equiv 0$, *Condition 2.1(c) is weaken as* $e^{2T|c(t,i,a)|} \le MV(i)$*)*

Compared to the risk-neutral (linear utility) case, to ensure the performance $V(\pi, i)$ to be finite in the risk-sensitive setup, it is necessary to impose more restrictive conditions on the growth of the cost rate. Part (c) of Condition 2.1 is motivated by part (b) of the next lemma and the Jensen inequality, see the proof of Lemma 2.1 below.

Lemma 2.1. *Suppose Condition 2.1 is satisfied. For each* $\pi \in \Pi$, *the following assertions hold.*

(a) $P_i^\pi(t_\infty = \infty) = 1$ *for each* $i \in \boldsymbol{S}$.

(b) $E_i^\pi[V(\xi_t)] \le e^{\rho t}V(i)$, *for each* $t \ge 0$ *and* $i \in \boldsymbol{S}$.

(c) $\mathcal{V}(\pi, i) \le Me^{T\rho}V(i)$ *for all* $i \in \boldsymbol{S}$ *and* $\pi \in \Pi$.

Proof. Parts (a) and (b) are known, see e.g., [54, 89, 90]. We next verify part (c). By part (a), for P_i^π-almost all $\omega \in \Omega$, there are finitely many values taken by in

$\{\xi_t(\omega)\}$ over $[0,T]$. For such $\omega \in \Omega$, by Condition 2.1(c), we legitimately write

$$\int_0^T \int_{\mathbf{A}} c(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_T) = \int_{(0,T]} \int_{\mathbf{A}} \tilde{c}(t,\xi_t,a)\pi(da|\omega,t)\mu(dt),$$

where $\mu(dt) = I_{[0,T)}(t)dt + \delta_T(dt)$, with $\delta_T(dt)$ being the Dirac measure concentrated on $\{T\}$, and $\tilde{c}(t,i,a) := c(t,i,a)I_{[0,T)}(t) + g(i)I_{\{T\}}(t)$ for each $(t,i,a) \in \mathbb{K}$. Now,

$$\begin{aligned}
& E_i^\pi \left[e^{\int_0^T \int_{\mathbf{A}} c(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_T)} \right] \\
= \; & E_i^\pi \left[e^{\int_{[0,T]} \int_{\mathbf{A}} (1+T)\tilde{c}(t,\xi_t,a)\pi(da|\omega,t) \frac{\mu(dt)}{T+1}} \right] \\
\leq \; & E_i^\pi \left[\frac{1}{1+T} \int_{[0,T]} e^{(1+T)\int_{\mathbf{A}} |\tilde{c}(t,\xi_t,a)|\pi(da|\omega,t)} \mu(dt) \right] \\
\leq \; & \frac{M}{1+T} E_i^\pi \left[\int_0^T V(\xi_t)dt + V(\xi_T) \right] \\
\leq \; & Me^{\rho T} V(i) \qquad (2.1)
\end{aligned}$$

where the first inequality is by the Jensen inequality, the second inequality is by Condition 2.1(c), and the last inequality is by part (b). □

Part (a) of the previous lemma asserts that under the imposed conditions therein, the controlled process is nonexplosive under each policy. This fact is used in the proof of Theorem 2.1 below, see the first paragraph therein as well as (2.7).

Condition 2.2. *There exist a $[1,\infty)$-valued function V_1 defined on $\boldsymbol{S}$, and constants $\rho_1 > 0$, $M_1 > 0$ such that*

(a) $\sum_{j \in \boldsymbol{S}} V_1^2(j) q(j|t,i,a) \leq \rho_1 V_1^2(i)$ for each $(t,i,a) \in \mathbb{K}$;

(b) $V^2(i) \leq M_1 V_1(i)$ for all $i \in \boldsymbol{S}$, with the function V as the Condition 2.1.

The role of this condition is seen in the proof of Theorem 2.1 below, where the Cauchy-Schwarz inequality is used, see (2.4) therein. Conditions 2.1 and 2.2 guarantee the growth of the value function and its derivative to be suitably

bounded by the drift functions V and V_1, and it is out of this class of functions that we show the Feynman-Kac formula applies. The previous works [43, 101] only showed that the Feynman-Kac formula is applicable to a class of bounded functions, and so confined themselves to the class of bounded cost rates, which excludes some potentially interesting applications. Let us formulate such an example, which are with unbounded transition and cost rates and satisfy Conditions 2.1 and 2.2.

Example 2.1. Consider a controlled $M/M/\infty$ queueing system, where the common service rate a of each server can be tuned from a finite interval $[\underline{\mu}, \overline{\mu}] \subseteq [0, \infty]$. Let the arrival rate be denoted by $\lambda > 0$. The holding cost is $C_1 i$ given the current number of jobs in the system being $i \geq 0$, where $C_1 > 0$ is a constant, and maintaining a service rate at μ costs μ per unit time. A terminal reward of $C_2 i$ is received if there are i jobs remaining in the system at the end of the horizon $[0, T]$, where $C_2 \in (-\infty, \infty)$ is a constant. The decision maker aims at the optimal control of the service rate to minimize the expected exponential utility of the total cost over the horizon $[0, T]$.

This problem can be formulated as a CTMDP with the following primitives. The state space is $\mathbf{S} = \{0, 1, \dots\}$, the action space is $[\underline{\mu}, \overline{\mu}] \equiv \mathbf{A}(t, i)$. The transition rate is given by $q(i+1|t, i, a) \equiv \lambda$, $q(i-1|t, i, a) = ai$ if $i \geq 1$, $q_i(a) = \lambda + ai$ if $i > 0$, and $q_0(a) = \lambda$. The running cost rate is given by $c(t, i, a) = C_1 i + a$, and the terminal cost is given by $g(i) = -C_2 i$.

Observe the following. Let $d > 0$ be a fixed constant. Let $\rho(d) := e^{d+1}\lambda$. Then for each constant $\rho \geq \rho(d)$, $\sum_{j \in \mathbf{S}} q(j|t, i, a) e^{dj} = e^{d(i+1)}\lambda + e^{d(i-1)}a - (\lambda + a)e^{di} \leq \rho e^{di}$ for each $i \geq 1$, and $\sum_{j \in \mathbf{S}} q(j|t, 0, a) e^{dj} = \lambda e^d - \lambda \leq \rho$. Therefore, for the verification of Condition 2.1, one can take $M = e^{2(1+T)\overline{\mu}} + \overline{\mu} + \lambda$, $V(i) = e^{d_1 i}$ with $d_1 = 2(1 + T)(C_1 + |C_2|)$, $\rho = \rho(d_1)$. For the verification of Condition 2.2, one can take $M_1 = 1$, and $V_1(i) = e^{d_2 i}$ with $d_2 = 2d_1$, and $\rho_1 = \rho(d_2)$.

Theorem 2.1. *Suppose Conditions 2.1 and 2.2 are satisfied. Then, for each*

$i \in \boldsymbol{S}$, $\pi \in \Pi$ *and* $\varphi \in C^1_{V,V_1}([0,T] \times \boldsymbol{S})$,

$$\begin{aligned} & E_i^\pi \left[\int_0^T \left(\psi'(\omega,t,\xi_t) + \sum_{j \in \boldsymbol{S}} \psi(\omega,t,j) \int_{\boldsymbol{A}} q(j|t,\xi_t,a)\pi(da|\omega,t) \right) dt \right] \\ = \; & E_i^\pi \left[\psi(\omega,T,\xi_T) \right] - \varphi(0,i), \end{aligned}$$

where outside a P_i^π*-null set, say* $\Omega \setminus \Omega'$, $T_\infty = \infty$,

$$\psi(\omega,t,j) = e^{\int_0^t \int_{\boldsymbol{A}} c(v,\xi_v,a)\pi(da|\omega,v)dv} \varphi(t,j), \;\; \forall \; t \in [0,T], \;\; j \in \boldsymbol{S},$$

$\psi(\omega,\cdot,j)$ *is absolutely continuous on* $[0,T]$ *so that we can take*

$$\begin{aligned} \psi'(\omega,t,j) \;\; = \;\; & \int_{\boldsymbol{A}} c(t,\xi_t,a)\pi(da|\omega,t) e^{\int_0^t \int_{\boldsymbol{A}} c(v,\xi_v,a)\pi(da|\omega,v)dv} \varphi(t,j) \\ & + e^{\int_0^t \int_{\boldsymbol{A}} c(v,\xi_v,a)\pi(da|\omega,v)dv} \varphi'(t,j), \end{aligned} \tag{2.2}$$

for each $\omega \in \Omega'$ *and* $j \in \boldsymbol{S}$.

Proof. According to Lemma 2.1(a), we concentrate on Ω' on which $T_\infty = \infty$, and hence (2.2) holds. Since $\varphi \in C^1_{V,V_1}([0,T] \times \mathbf{S})$, we have $|\varphi(t,i)| \le \|\varphi\|_V V(i)$ for all $(t,i) \in [0,T] \times \mathbf{S}$, which, together with the relation $(1+T)|c(v,i,a)| \le MV(i)$ (by Condition 2.1(c)), leads to

$$\begin{aligned} & |\psi'(\omega,t,\xi_t)| \\ \le \; & \frac{M}{1+T} V(\xi_t) e^{\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)|\pi(da|\omega,v)dv} \|\varphi\|_V V(\xi_t) + \|\varphi'\|_{V_1} e^{\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)|\pi(da|\omega,v)dv} V_1(\xi_t), \\ \le \; & \frac{\|\varphi\|_V + \|\varphi'\|_{V_1}}{1+T} (1 + T + MM_1) e^{\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)|\pi(da|\omega,v)dv} V_1(\xi_t). \end{aligned} \tag{2.3}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} & E_i^\pi \left[e^{\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)|\pi(da|\omega,v)dv} V_1(\xi_t) \right] \le \sqrt{E_i^\pi \left[e^{2\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)|\pi(da|\omega,v)dv} \right] E_i^\pi \left[V_1^2(\xi_t) \right]} \\ \le \; & E_i^\pi \left[e^{2\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)|\pi(da|\omega,v)dv} \right] E_i^\pi \left[V_1^2(\xi_t) \right] \le M e^{T\rho} V(i) E_i^\pi \left[V_1^2(\xi_t) \right] \\ \le \; & M e^{T\rho} V(i) e^{\rho_1 T} V_1^2(i), \;\; t \in [0,T], \end{aligned} \tag{2.4}$$

where the second to the last inequality is obtained by a similar argument to the one for (2.1), and the last inequality is by Lemma 2.1(b). Now it follows from (2.3) that

$$E_i^\pi \left[\int_0^T |\psi'(\omega, t, \xi_t)| dt \right] < \infty. \tag{2.5}$$

On the other hand, by Conditions 2.1 and 2.2, we have

$$\begin{aligned} & \sum_{j \in \mathbf{S}} e^{\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)| \pi(da|\omega,v) dv} |\varphi(t,j)| \left| \int_{\mathbf{A}} q(j|t,\xi_t,a) \pi(da|\omega,t) \right| \\ \leq & \ \|\varphi\|_V \left(\rho V(\xi_t) + 2MV^2(\xi_t) \right) e^{\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)| \pi(da|\omega,v) dv} \\ \leq & \ \|\varphi\|_V M_1(\rho + 2M) e^{\int_0^t \int_{\mathbf{A}} |c(v,\xi_v,a)| \pi(da|\omega,v) dv} V_1(\xi_t). \end{aligned}$$

Now it follows from (2.4) that

$$\int_0^T \sum_{j \in \mathbf{S}} E_i^\pi \left[\left| \int_{\mathbf{A}} q(j|t,\xi_t,a) \pi(da|\omega,t) \right| |\psi(\omega,t,j)| \right] dt < \infty. \tag{2.6}$$

For each $0 \leq s \leq T$,

$$\psi(\omega, T, \xi_T) = \psi(\omega, 0, \xi_0) + \int_0^T \psi'(\omega, t, \xi_t) dt + \sum_{n \geq 1} \int_{(0,T]} \Delta\psi(\omega, t, \xi_t) \delta_{T_n}(dt) \tag{2.7}$$

with $\Delta\psi(\omega, t, \xi_t) := \psi(\omega, t, \xi_t) - \psi(\omega, t-, \xi_{t-})$. (Recall that the function $\psi(\omega, t, j)$ is absolutely continuous in t over any finite interval, and for each fixed $\omega \in \Omega'$ with Ω' being defined in the beginning of this proof, $\xi_t(\omega)$ is piecewise constant in $t \in [0, T]$, and thus has finitely many values over that interval.) By (2.5) and (2.6), we take legitimately the expectation on the both sides of the previous

equality, and obtain

$$\begin{aligned} E_i^\pi\left[\psi(\omega,T,\xi_T)\right] &= E_i^\pi\left[\psi(\omega,0,\xi_0)\right]+E_i^\pi\left[\int_0^T \psi'(\omega,t,\xi_t)dt\right] \\ &\quad +E_i^\pi\left[\sum_{n\geq 1}\int_{(0,T]}\Delta\psi(\omega,t,\xi_t)\delta_{T_n}(dt)\right] \\ &= \varphi(0,i)+E_i^\pi\left[\int_0^T \psi'(\omega,t,\xi_t)dt\right] \\ &\quad +E_i^\pi\left[\sum_{j\in\mathbf{S}}\int_{(0,T]}(\psi(\omega,t,j)-\psi(\omega,t,\xi_{t-}))m^\pi(j|\omega,t)dt\right] \\ &= \varphi(0,i)+E_i^\pi\left[\int_0^T \psi'(\omega,t,\xi_t)dt\right] \\ &\quad +E_i^\pi\left[\sum_{j\in\mathbf{S}}\int_0^T\int_{\mathbf{A}}\psi(\omega,t,j)q(j|t,\xi_{t-},a)\pi(da|\omega,t)dt\right], \end{aligned}$$

where the last equality holds because the random measure m^π is the dual predictable projection of the random measure $\sum_{n\geq 1}\delta_{(T_n,X_n)}(dt,dx)$ on $\mathcal{B}((0,\infty)\times\mathbf{S})$ under P_i^π, see p.131 of [71]. The statement is proved. □

The above Feynman-Kac formula in the above theorem was justified in [101], see Theorem 3.1 therein, when π is a Markov policy, and φ is assumed to be bounded.

The next statement provides a verification theorem, which was known in [88] when the transition rate is bounded.

Corollary 2.1. *Suppose Conditions 2.1 and 2.2 are satisfied. If there exists* $\varphi\in C^1_{V,V_1}([0,T]\times\boldsymbol{S})$ *and a deterministic Markov policy* $f\in\Pi_m^d$ *such that*

$$\begin{aligned} \varphi(s,i)-e^{g(i)} &= \int_s^T \inf_{a\in\boldsymbol{A}(t,i)}\left\{c(t,i,a)\varphi(t,i)+\sum_{j\in\boldsymbol{S}}\varphi(t,j)q(j|t,i,a)\right\}dt \\ &= \int_s^T\left\{c(t,i,f(t,i))\varphi(t,i)+\sum_{j\in\boldsymbol{S}}\varphi(t,j)q(j|t,i,f(t,i))\right\}dt, \\ &\quad s\in[0,T],\ i\in\boldsymbol{S}, \end{aligned} \tag{2.8}$$

then

$$\mathcal{V}(f,i)=\varphi(0,i)=\mathcal{V}^*(i),\ \forall\ i\in \boldsymbol{S}. \tag{2.9}$$

Proof. Concentrate on Ω' as in the proof of the previous theorem. It holds for almost all $t\in[0,T]$ that

$$\begin{aligned}
& 0=\varphi'(t,\xi_t)+\inf_{a\in\mathbf{A}(t,\xi_t)}\left\{c(t,\xi_t,a)\varphi(t,\xi_t)+\sum_{j\in\mathbf{S}}\varphi(t,j)q(j|t,\xi_t,a)\right\}\\
= \ & \varphi'(t,\xi_t)+c(t,\xi_t,f(t,\xi_t))\varphi(t,\xi_t)+\sum_{j\in\mathbf{S}}\varphi(t,j)q(j|t,\xi_t,f(t,\xi_t))\\
\leq \ & \varphi'(t,\xi_t)+\int_{\mathbf{A}}\left\{c(t,\xi_t,a)\varphi(t,\xi_t)+\sum_{j\in\mathbf{S}}\varphi(t,j)q(j|t,\xi_t,a)\right\}\pi(da|\omega,t).
\end{aligned}$$

Now by applying Theorem 2.1 to the deterministic Markov policy f and an arbitrarily fixed $\pi\in\Pi$, we see

$$\begin{aligned}
\mathcal{V}(\pi,i)-\varphi(0,i) &= E_i^{\pi}\left[e^{\int_0^T\int_{\mathbf{A}}c(v,\xi_v,a)\pi(da|\omega,v)dv}\varphi(T,\xi_T)\right]-\varphi(0,i)\\
&= E_i^{\pi}\left[\int_0^T e^{\int_0^t\int_{\mathbf{A}}c(v,\xi_v,a)\pi(da|\omega,v)dv}\int_{\mathbf{A}}\{c(t,\xi_t,a)\varphi(t,\xi_t)\right.\\
&+ \left.\varphi'(t,\xi_t)+\sum_{j\in\mathbf{S}}\varphi(t,j)q(j|t,\xi_t,a)\right\}\pi(da|\omega,t)\Bigg]\\
&\geq 0,
\end{aligned}$$

where the first equality holds because $\varphi(T,i)=e^{g(i)}$, see (2.8); similarly, replacing f for π in the equalities in the above, $\mathcal{V}(f,i)-\varphi(0,i)=0$. Consequently, $\mathcal{V}(f,i)=\varphi(0,i)\leq\mathcal{V}(\pi,i)$ for each $i\in\mathbf{S}$. Since π was arbitrarily fixed, $\mathcal{V}(f,i)=\varphi(0,i)=\mathcal{V}^*(i)$, as required. □

According to the previous statement, (2.8) is called the optimality equation, and the policy f in (2.9) is optimal.

The next statement was basically obtained in Theorem 2.1 in [43], see also [101].

Proposition 2.1. *Suppose that the transition and cost rates are bounded, i.e.,*

$$\sup_{i\in\boldsymbol{S}} \bar{q}_i < \infty, \quad \sup_{(t,i,a)\in\mathbb{K}} |c(t,i,a)| < \infty, \quad \sup_{i\in\boldsymbol{S}} |g(i)| < \infty.$$

If for each $i \in \boldsymbol{S}$ and $t \in [0,T]$, $\boldsymbol{A}(t,i)$ is compact, $c(t,i,a)$ is lower semicontinuous in $a \in \boldsymbol{A}(t,i)$, and $q(j|t,i,a)$ is continuous in $a \in \boldsymbol{A}(t,i)$, then there exists a unique φ in $C^1_{1,1}([0,T]\times\boldsymbol{S})$ and some $f \in \Pi^d_m$ satisfying (2.8) and (2.9).

The main objective in this chapter is to relax the boundedness requirements in the previous statement.

2.3 Optimality results

We impose the following condition, which guarantees the existence of an optimal policy.

Condition 2.3. *(a) For each $t \in [0,T], i,j \in \boldsymbol{S}$, the function $q(j|t,i,a)$ is continuous in $a \in \boldsymbol{A}(t,i)$.*

(b) For each $(t,i) \in [0,T]\times\boldsymbol{S}$, the function $c(t,i,a)$ is lower semicontinuous in $a \in \boldsymbol{A}(t,i)$, and the function $\sum_{j\in\boldsymbol{S}} V(j)q(j|t,i,a)$ is continuous in $a \in \boldsymbol{A}(t,i)$, with V as in Condition 2.1.

Under Conditions 2.1 and 2.3, the function $\sum_{j\in\mathbf{S}} q(j|t,i,a)u(t,j)$ is continuous in $a \in \mathbf{A}(t,i)$, for every fixed $(t,i) \in [0,T]\times\mathbf{S}$ and V-bounded measurable function u on $[0,T]\times\mathbf{S}$, see the proof of Lemma 8.3.7(a) in [58]. This fact will be used in the proof of the next statement.

Also note that Condition 2.3 is satisfied by Example 2.1.

The main optimality result is the following one.

Theorem 2.2. *Suppose Conditions 2.1, 2.2 and 2.3 are satisfied. Then there exists a unique φ in $C^1_{V,V_1}([0,T]\times\boldsymbol{S})$ and some $f \in \Pi^d_m$ satisfying (2.8) and (2.9). In particular, there exists a deterministic Markov optimal policy.*

Proof. The statement would follow from Corollary 2.1, once we showed the existence of some $\varphi \in C^1_{V,V_1}([0,T]\times \mathbf{S})$ satisfying (2.8). We verify this fact following a similar reasoning as in [52] dealing with a risk-neutral CTMDP problem, which was also adopted in [101], dealing with a model with a bounded cost rate. Namely, we shall obtain the desired solution φ as a limit point of an equicontinuous family $\{\varphi_n\}$ of functions, which in turn are obtained from a sequence of CTMDP models with bounded transition and cost rates. The denumerable state space serves to prove the equicontinuity of the family $\{\varphi_n\}$. The details are as follows.

For each integer $n \geq 1$, let $\mathbf{S}_n := \{i \in \mathbf{S} : \ V(i) \leq n\}$. Without loss of generality, assume for each $n \geq 1$, $\mathbf{S}_n \neq \emptyset$. For each $i \in \mathbf{S}$ and $t \in [0,\infty)$, let $\mathbf{A}_n(t,i) := \mathbf{A}(t,i)$. For each $(t,i,a) \in \mathbb{K}_n := \mathbb{K}$, define

$$
\begin{aligned}
&q_n(j|t,i,a) := q(j|t,i,a)I_{\mathbf{S}_n}(i),\ \forall\ j \in \mathbf{S}\\
&c_n(t,i,a) := c(t,i,a)I_{\mathbf{S}_n}(i),\ \ g_n(i) := g(i)I_{\mathbf{S}_n}(i).
\end{aligned}
$$

We consider the resulting sequence of CTMDP models $\mathcal{M}_n := \{\mathbf{S}, \mathbf{A}_n(t,i), c_n, g_n, q_n\}$.

Note that the models $\{\mathcal{M}_n\}$ are all with bounded transition and cost rates, and so Proposition 2.1 implies, for each $n \geq 1$, the existence of a unique φ_n in $C^1_{1,1}([0,T]\times\mathbf{S})$ and some $f_n \in \Pi^d_m$ satisfying

$$
\begin{aligned}
\varphi_n(s,i) - e^{g_n(i)} &= \int_s^T \inf_{a\in\mathbf{A}(t,i)} \left\{ c_n(t,i,a)\varphi_n(t,i) + \sum_{j\in\mathbf{S}} \varphi_n(t,j) q_n(j|t,i,a) \right\} dt\\
&= \int_s^T \left\{ c_n(t,i,f_n(t,i))\varphi_n(t,i) + \sum_{j\in\mathbf{S}} \varphi_n(t,j) q_n(j|t,i,f_n(t,i)) \right\} dt,\\
&\quad s \in [0,T],\ i \in \mathbf{S}. \qquad (2.10)
\end{aligned}
$$

Let $n \geq 1$ be fixed. For each $s \in [0,T]$, consider the s-shifted model

$$
\mathcal{M}_n^{(s)} := \left\{ \mathbf{S}, \mathbf{A}_n^{(s)}(t,i), q_n^{(s)}, c_n^{(s)}, g_n \right\}
$$

with $\mathbf{A}_n^{(s)}(t,i) := \mathbf{A}_n(t+s,i)$, $q_n^{(s)}(\cdot|t,i,a) := q_n(\cdot|s+t,i,a)$ and $c_n^{(s)}(t,i,a) :=$

$c_n(t+s,i,a)$. Then Condition 2.1 is clearly satisfied by $\mathcal{M}_n^{(s)}$, so that one can apply the reasoning in the proof of Lemma 2.1(c) and deduce

$$\mathrm{E}_i^{f_n^{(s)}}\left[e^{\int_0^{T-s}|c_n^{(s)}(t,\xi_t,f_n^{(s)}(t,\xi_t))|dt+|g_n(\xi_{T-s})|}\right]\leq Me^{T\rho}V(i)$$

where $\mathrm{E}_i^{f_n^{(s)}}$ denotes the expectation in the $\mathcal{M}_n^{(s)}$ model under the shifted policy $f_n^{(s)}(t,i):=f_n(t+s,i)$. On the other hand, according to the uniqueness of the solution to (2.10) in $C_{1,1}^1([0,T]\times\mathbf{S})$ and the second application of main theorem in section 5.1 of chapter 5,

$$\mathrm{E}_i^{f_n^{(s)}}\left[e^{\int_0^{T-s}c_n^{(s)}(t,\xi_t,f_n^{(s)}(t,\xi_t))dt+g_n(\xi_{T-s})}\right]=\varphi_n(s,i).$$

(The cost rate and the terminal cost were assumed to be nonnegative in chapter 5, but the results obtained there apply because $\mathcal{M}_n^{(s)}$ has bounded transition and cost rates, which can be reduced to the nonnegative case after one add to the cost rate and the terminal cost a large enough constant.) Thus, we obtain the bound

$$|\varphi_n(t,i)|\leq Me^{T\rho}V(i),\ \forall\ n\geq 1,(t,i)\in[0,T]\times\mathbf{S}. \tag{2.11}$$

which means $|\varphi_n(t,i)|$ is uniformly bounded.

Next, we show that $\{\varphi_n,n\geq 1\}$ is an equicontinuous family of functions on $[0,T]\times\mathbf{S}$, as follows. Let

$$H_n(t,i):=\inf_{a\in\mathbf{A}_n(t,i)}\left\{c_n(t,i,a)\varphi_n(t,i)+\sum_{j\in\mathbf{S}}\varphi_n(t,j)q_n(j|t,i,a)\right\},\ \forall\ (t,i)\in[0,T]\times\mathbf{S}.$$

Then, from Condition 2.1 and (2.11), we see

$$\begin{aligned}
|H_n(t,i)| &\leq \sup_{a\in\mathbf{A}_n(t,i)}\left\{|c_n(t,i,a)\varphi_n(t,i)|+\sum_{j\in\mathbf{S}}|\varphi_n(t,j)||q_n(j|t,i,a)|\right\}\\
&\leq \sup_{a\in\mathbf{A}_n(t,i)}\left\{MV(i)Me^{T\rho}V(i)+Me^{T\rho}\sum_{j\in\mathbf{S}}|q(j|t,i,a)|V(j)\right\}\\
&\leq e^{T\rho}(M^2V^2(i)+\rho MV(i)+2M|q(i|t,i,a)|V(i))\\
&\leq Me^{T\rho}M_1(3M^2+\rho)V_1(i)=:L(i),\ \forall\ (t,i)\in[0,T]\times\mathbf{S}. \quad (2.12)
\end{aligned}$$

(Recall that $M>1$.)

Now, fix arbitrarily some $(s_0,i_0)\in[0,T]\times\mathbf{S}$ and $\varepsilon>0$, and take $\delta:=\min\{\frac{\varepsilon}{L(i_0)},\frac{1}{2}\}$. Then, for every (s,i) in the open neighborhood $\{(s,i)\in[0,T]\times\mathbf{S}:\ |s-s_0|<\delta,|i-i_0|<\delta\}$, we have $i=i_0$, and

$$\begin{aligned}
|\varphi_n(s,i)-\varphi_n(s_0,i_0)| &= |\varphi_n(s,i_0)-\varphi_n(s_0,i_0)|=\left|\int_s^T H_n(t,i_0)dt-\int_{s_0}^T H_n(t,i_0)dt\right|\\
&\leq L(i_0)|s-s_0|<\varepsilon,\ \forall\ n\geq 1.
\end{aligned}$$

Hence, $\{\varphi_n,n\geq 1\}$ is equicontinuous at (s_0,i_0), which, together with the arbitrariness of $(s_0,i_0)\in[0,T]\times\mathbf{S}$, yields that $\{\varphi_n,n\geq 1\}$ is equicontinuous on $[0,T]\times\mathbf{S}$. By Arzela-Ascoli theorem, see, e.g., p.96 of [57], there exist a subsequence $\{\varphi_{n_k},k\geq 1\}$ of $\{\varphi_n,n\geq 1\}$ and a continuous function φ on $[0,T]\times\mathbf{S}$ such that

$$\lim_{k\to\infty}\varphi_{n_k}(s,i)=\varphi(s,i),\ \text{and}\ |\varphi(s,i)|\leq Me^{T\rho}V(i)\ \ \forall\ (s,i)\in[0,T]\times\mathbf{S},\ (2.13)$$

where the last inequality is by (2.11).

Let

$$H(t,i):=\inf_{a\in\mathbf{A}(t,i)}\left\{c(t,i,a)\varphi(t,i)+\sum_{j\in\mathbf{S}}\varphi(t,j)q(j|t,i,a)\right\},\forall\ (t,i)\in[0,T]\times\mathbf{S}.$$

We next verify that $\lim_{k\to\infty}H_{n_k}(t,i)=H(t,i)$ for each $(t,i)\in[0,T]\times\mathbf{S}$, as

follows. Let $(t,i) \in [0,T] \times \mathbf{S}$ be arbitrarily fixed. Since $q_{n_k}(j|t,i,a) \to q(j|t,i,a)$ for all $j \in \mathbf{S}$ and $a \in \mathbf{A}(t,i)$ as $k \to \infty$, by virtue of Lemma 8.3.7 in [58] and (2.11), we have

$$\begin{aligned} \limsup_{k\to\infty} H_{n_k}(t,i) &\leq \limsup_{k\to\infty} \left\{ c_{n_k}(t,i,a)\varphi_{n_k}(t,i) + \sum_{j\in\mathbf{S}} \varphi_{n_k}(t,j) q_{n_k}(j|t,i,a) \right\} \\ &\leq c(t,i,a)\varphi(t,i) + \sum_{j\in\mathbf{S}} \varphi(t,j) q(j|t,i,a), \ \forall\ a \in \mathbf{A}(t,i), \end{aligned}$$

so that

$$\limsup_{k\to\infty} H_{n_k}(t,i) \leq \inf_{a\in\mathbf{A}(t,i)} \left\{ c(t,i,a)\varphi(t,i) + \sum_{j\in\mathbf{S}} \varphi(t,j) q(j|t,i,a) \right\}. \quad (2.14)$$

According to the fact mentioned below Condition 2.3, there exists a sequence of policies $\{f_{n_k}\} \subseteq \Pi_m^d$ such that

$$\begin{aligned} H_{n_k}(t,i) &= \inf_{a\in\mathbf{A}(t,i)} \left\{ c_{n_k}(t,i,a)\varphi_{n_k}(s,i) + \sum_{j\in\mathbf{S}} \varphi_{n_k}(t,j) q_{n_k}(j|t,i,a) \right\} \\ &= c(t,i,f_{n_k}(t,i))\varphi_{n_k}(t,i) + \sum_{j\in\mathbf{S}} \varphi_{n_k}(t,j) q_{n_k}(j|t,i,f_{n_k}(t,i)). \end{aligned}$$

Since $\mathbf{A}(t,i)$ is compact, by taking subsequences if necessary, we can assume without loss of generality that $\liminf_{k\to\infty} H_{n_k}(t,i) = \lim_{k\to\infty} H_{n_k}(t,i)$ and for some $a \in \mathbf{A}(t,i)$, $f_{n_k}(t,i) \to a$ as $k \to \infty$. By the virtue of Lemma 8.3.7 in [58], we have

$$\begin{aligned} &\liminf_{k\to\infty} H_{n_k}(t,i) = \liminf_{k\to\infty} \left\{ c(t,i,f_{n_k}(t,i))\varphi_{n_k}(t,i) + \sum_{j\in\mathbf{S}} \varphi_{n_k}(t,j) q_{n_k}(j|t,i,f_{n_k}(t,i)) \right\} \\ \geq\ & c(s,i,a)\varphi(t,i) + \sum_{j\in\mathbf{S}} \varphi(t,j) q(j|t,i,a) \geq \inf_{a\in\mathbf{A}(t,i)} \left\{ c(t,i,a)\varphi(t,i) + \sum_{j\in\mathbf{S}} \varphi(t,j) q(j|t,i,a) \right\}. \end{aligned}$$

(Recall Condition 2.3.) This, together with (2.14), implies that $\lim_{k\to\infty} H_{n_k}(t,i) = H(t,i)$. Since $(t,i) \in [0,T] \times \mathbf{S}$ was arbitrarily fixed, we see from (2.10), (2.12)

and (2.13) that φ satisfies (2.8). The same argument as in (2.12) leads to

$$|\varphi'(t,i)| = |H(t,i)| \le Me^{T\rho}M_1(3M^2+\rho)V_1(i),\ \forall\ (t,i) \in [0,T]\times \mathbf{S}.$$

Therefore, we see that $\varphi \in C^1_{V,V_1}([0,T]\times\mathbf{S})$. The required deterministic Markov policy f exists because of the fact mentioned below Condition 2.3, a measurable selection theorem, see Proposition D.5 of [57].

Finally, we verify the uniqueness part. Let $\varphi \in C^1_{V,V_1}([0,T]\times\mathbf{S})$ be an arbitrarily fixed solution to (2.8). (The above reasoning shows that there exists at least one.) Let $s \in [0,T]$ be fixed, and consider the s-shifted model $\mathcal{M}^{(s)} = \left\{\mathbf{S}, \mathbf{A}^{(s)}(t,i), q^{(s)}, c^{(s)}, g\right\}$, which is defined as for the $\mathcal{M}^{(s)}_n$ model with n being omitted everywhere. Let

$$\mathcal{V}^{(s)}(i) := \inf_{\pi\in\Pi} \mathrm{E}^{\pi}_i\left[e^{\int_0^{T-s}\int_{\mathbf{A}} c^{(s)}(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_{T-s})}\right]$$

with E^{π}_i signifying the expectation in the s-shifted model. Then the function $\varphi^{(s)} \in C^1_{V,V_1}([0,T-s]\times\mathbf{S})$ defined by $\varphi^{(s)}(\tau,i) := \varphi(\tau+s,i)$ for each $(\tau,i) \in [0,T-s]\times\mathbf{S}$ satisfies

$$\begin{aligned}
\varphi^{(s)}(\tau,i) - e^{g(i)} &= \int_\tau^{T-s} \inf_{a\in\mathbf{A}^{(s)}(t,i)} \left\{c^{(s)}(t,i,a)\varphi^{(s)}(t,i) + \sum_{j\in\mathbf{S}}\varphi^{(s)}(t,j)q^{(s)}(j|t,i,a)\right\}dt \\
&= \int_\tau^{T-s} \left\{c^{(s)}(t,i,f^{(s)}(t,i))\varphi^{(s)}(t,i) + \sum_{j\in\mathbf{S}}\varphi^{(s)}(t,j)q^{(s)}(j|t,i,f^{(s)}(t,i))\right\}dt, \\
&\quad \tau\in[0,T-s],\ i\in\mathbf{S},
\end{aligned}$$

for some deterministic Markov policy $f^{(s)}$. By applying Corollary 2.1 to the s-shifted model $\mathcal{M}^{(s)}$, we see $\varphi^{(s)}(0,i) = \mathcal{V}^{(s)}(i)$, and thus $\varphi(s,i) = \mathcal{V}^{(s)}(i)$ for each $i \in \mathbf{S}$. Since $s \in [0,T]$ was arbitrarily fixed, it follows that φ is the unique solution to (2.8) out of $\varphi \in C^1_{V,V_1}([0,T]\times\mathbf{S})$. The proof is completed. □

Remark 2.1. We can refer to Chapter 6 for risk-sensitive piecewise deterministic Markov decision processes (RS-PDMDP), where we get the optimality results for RS-PDMDP with non-negative cost rates by the technique of reducing the

original problem to a RS-DTMDP problem. As an application, finite horizon and infinite horizon discounted RS-CTMDP can be reformulated as total undiscounted RS-PDMDP. As we can notice naturally, if the cost rate in this chapter can transfer from the drift function bounded to nonnegative, then we can easily use the conclusion of PDMDP in this chapter to get the results, but this is still an open problem.

I have to mention that recently we find there is an outstanding research published in 2019 that can not only cover our results but extend it to a more general case, see [63]. It deals with finite horizon RS-PDMDP with reward rates bounded by drift function (need not to be nonnegative), and the state space is Borel space. Compared with ours, it can be used more extensively. [63] adapts the approach where the value function is characterized as a solution to the related integro-differential HJB equation. And it develops Feyman Kac's formula for PDMDPs with unbounded transition rates.

3 Risk-sensitive average CTMDP with unbounded rates

Having the results of finite horizon risk-sensitive CTMDP with unbounded cost and transition rates, we can keep on to get the average case problem based on them. Because in the risk-sensitive average CTMDP problem, we always consider the finite horizon case first. This chapter considers the *risk-sensitive average optimization* for denumerable CTMDPs, in which the transition and cost rates are allowed to be unbounded, and the policies can be randomized history-dependent. We first derive the multiplicative dynamic programming principle and some new properties for the risk-sensitive finite-horizon CTMDPs. Then, we establish the existence and uniqueness of a solution to the risk-sensitive average optimality equation (RS-AOE) by the results for risk-sensitive finite-horizon CTMDPs developed here, and also prove the existence of an optimal stationary policy via the RS-AOE and the extended Feymanm-Kac's formula. Furthermore, for the case of finite actions available at each state, we construct a sequence of models of finite-state CTMDPs with optimal stationary policies which can be obtained by a policy iteration algorithm in a finite number of iterations, and prove that an average optimal policy for the case of infinitely countable states can be approximated by those of the finite-state models. Finally, we illustrate the conditions in this paper and show the difference between the conditions here and those in the previous literature with some examples.

3.1 On the risk-sensitive finite-horizon optimality

Here, the state space is also denumerable but the model is homogeneous. We assume $c(i,a)$ is bounded below (i.e., $c(i,a) \geq L$ for all $(i,a) \in \mathbb{K}$, for some

constant L), so we have $c(i,a)+|L|\geq 0$ on $\mathbb{K}$, and

$$\limsup_{T\to\infty}\frac{1}{T}\ln E_i^{\pi}\left[e^{\int_0^T\int_A[c(\xi_t,a)+|L|]\pi(da|\omega,t)dt}\right]=J(i,\pi)+|L|$$

Thus, adding a constant to all the costs $c(i,a)$ will affect all policies identically in both criteria, we may without loss of generalization assume that the costs $c(i,a)$ are *nonnegative*.

To prove the existence of risk-sensitive average optimal policies, we need to develop some preliminary facts about the risk-sensitive finite-horizon CTMDPs, some of which are from Chapter 2, and some of which are new.

Since the transition and cost rates (i.e., $q(j|i,a)$ and $c(i,a)$) may be unbounded, to guarantee the non-explosion of $\{\xi_t, t\geq 0\}$ and the finiteness of $J(i,\pi)$, we need the following conditions from [47, 52, 54, 74]

Condition 3.1. *There exist real-valued functions $V_0\geq 1$ and $\delta>0$ on $\boldsymbol{S}$, positive constants b_0 and M_0, and a state $i_0\in\boldsymbol{S}$, such that*

(a) $\sum_{j\in\boldsymbol{S}}V_0(j)q(j|i,a)\leq -\delta(i)V_0(i)+b_0I_{\{i_0\}}(i)$ for all $(i,a)\in\mathbb{K}:=\{(i,a)|i\in \boldsymbol{s},a\in\boldsymbol{A}(i)\}$;

(b) $\bar{q}_i\leq M_0V_0(i)$ *for all* $i\in\boldsymbol{S}$;

(c) $\delta_*:=\inf_{i\neq i_0}\delta(i)>0$, *and* $c(i,a)\leq\delta(i)\leq\sqrt{\ln V_0(i)}$ *for all* $(i,a)\in\mathbb{K}$.

Remark 3.1. (a) Although the indicator function $I_{\{i_0\}}$ in Condition 3.1(a) is stronger than the indicator function I_C in [74] with a finite subset C of $\mathbf{S}$, we will understand that such a restrictiveness is required; see Remark 3.4 below for detail. Condition 3.1(a) is an extension of Assumption (A2) in [74], Assumption (A5) in [43], and Assumption 7.1 in [50] from a constant δ to a function $\delta(i)$ on $\mathbf{S}$ here. Thus, it is satisfied for the examples in [43, 50] and will be verified with other examples below.

(b) Condition 3.1(c) is new and serves the finiteness of $E_i^{\pi}\left[e^{\int_0^{\tau_{i_0}}\int_{\mathbf{A}}c(\xi_v,a)\pi(da|\omega,v)dv}\right]$, where $\tau_{i_0}:=\min\{t\geq 0|\xi_t=i_0\}$ denotes the first passage time to i_0. Since the

cost $c(i,a)$ satisfying Condition 3.1(c) is allowed to be unbounded (see Example 3.2 and Remark 3.9 below), Condition 3.1(c) is weaker than the small condition (i.e., $\|c\| < \delta_*$) in [43, 74].

(c) Condition 3.1(a) is slightly stronger than Condition 2.1(a).

Lemma 3.1. *Under Conditions 3.1(a)(b), the following assertions hold.*

(a) $P_i^\pi(t_\infty = \infty) = 1$, *and* $P_i^\pi(\xi_t \in \boldsymbol{S}) = 1$ *for each* $t \geq 0$, $i \in \boldsymbol{S}$, *and* $\pi \in \Pi$.

(b) $E_i^\pi[V_0(\xi_t)] \leq V_0(i) + b_0 t$, *for each* $t \geq 0$, $i \in \boldsymbol{S}$, *and* $\pi \in \Pi$;

(c) $E_\gamma^\pi[V_0(\xi_t)\big|\xi_s = i] \leq V_0(i) + (t-s)b_0$, *for each* $t \geq s \geq 0$, $i \in \boldsymbol{S}$ *and* $\pi \in \Pi_m^r$.

Proof. (a) and (b) follow from Theorem 3.1 in [55], while (c) from Lemma 6.3 in [50]. □

Lemma 3.1 gives conditions for the non-explosion of $\{\xi_t, t \geq 0\}$ and also provides an estimate of $E_i^\pi[V_0(\xi_t)]$. In order to deal with the risk-sensitive average optimality, we next need some notations and facts on the risk-sensitive finite-horizon optimality.

For any $\pi \in \Pi, t \geq 0, i \in \mathbf{S}$, the following t-horizon risk-sensitive criterion

$$\varphi(t,i,\pi) := E_i^\pi\left[e^{\int_0^t \int_{\mathbf{A}(x_s)} c(\xi_s,a)\pi(da|\omega,s)ds}\right], \tag{3.1}$$

is well defined. Then, let

$$\varphi(t,i) := \inf_{\pi\in\Pi} \varphi(t,i,\pi) \quad (\text{for } i \in \mathbf{S}), \tag{3.2}$$

which is called the value function of the t-horizon risk-sensitive criterion. Since c is nonnegative, $\varphi(t,i)$ is increasing in $t \geq 0$, and $\varphi(0,i) = 1$ as well as $\varphi(t,i) \geq 1$ for all $t \geq 0$.

To further characterize risk-sensitive finite-horizon CTMDPs, we need the extension of Feymanm-Kac's formula in Theorem 2.1 from a Markov chain case

to a more general case of non-Markov processes, which is based on the following condition (similar as Condition 2.2).

Condition 3.2. *There exist a real-valued function $V_1 \geq 1$ on $\boldsymbol{S}$, and positive constants ρ_1, b_1, and M_1, such that*

(i) $\sum_{j\in\boldsymbol{S}} V_1^2(j)q(j|i,a) \leq \rho_1 V_1^2(i) + b_1$ for all $(i,a) \in \mathbb{K}$;

(ii) $V_0^2(i) \leq M_1 V_1(i)$ *for all* $i \in \boldsymbol{S}$, *where* V_0 *comes from Condition 3.1.*

Lemma 3.2. *(The extension of Feymanm-Kac's formula) Under Conditions 3.1 and 3.2, for any $T > 0$, the following assertions hold.*

(a) *For any $\pi \in \Pi$ and $u \in C^1_{V_0,V_1}([0,T] \times \boldsymbol{S})$,*

$$\begin{aligned} & E_i^\pi \Bigg[\int_0^{T\wedge\tau_{\boldsymbol{D}}} \Bigg(\Big(e^{\int_0^t \int_{\boldsymbol{A}} c(\xi_v,a)\pi(da|\omega,v)dv} u(t,\xi_t) \Big)' \\ & + \sum_{j\in\boldsymbol{S}} e^{\int_0^t \int_{\boldsymbol{A}} c(\xi_v,a)\pi(da|\omega,v)dv} u(t,j) \int_{\boldsymbol{A}} q(j|\xi_t,a)\pi(da|\omega,t) \Bigg) dt \Bigg] \\ = \; & E_i^\pi \left[e^{\int_0^{T\wedge\tau_{\boldsymbol{D}}} \int_{\boldsymbol{A}} c(\xi_v,a)\pi(da|\omega,v)dv} u(T,\xi_{T\wedge\tau_{\boldsymbol{D}}}) \right] - u(0,i), \quad i \in \boldsymbol{S} \end{aligned}$$

where $\tau_{\boldsymbol{D}} := \inf\{s \geq 0 | \xi_s \in \boldsymbol{D}\}$ is the hitting time of the process $\{\xi_t\}$ to a set $\boldsymbol{D} \subseteq \boldsymbol{S}$, and $\{\xi_t, t \geq 0\}$ may be not Markovian since the policy π may depend on histories.

(b) *For each $\pi = \pi_t(da|\cdot) \in \Pi_m^r$, and $u \in C^1_{V_0,V_1}([0,T] \times \boldsymbol{S})$,*

$$E_\gamma^\pi \left[\int_s^{T\wedge\tau_{\boldsymbol{D}}} \left(\left(e^{\int_s^t c(\xi_v,\pi_v)dv} u(t,\xi_t) \right)' + \sum_{j\in\boldsymbol{S}} \left(e^{\int_s^t c(\xi_v,\pi_v)dv} u(t,j) \right) q(j|\xi_t,\pi_t) \right) dt \Big| \xi_s = i \right]$$

$$= E_\gamma^\pi \left[e^{\int_s^{T\wedge\tau_{\boldsymbol{D}}} c(\xi_v,\pi_v)dv} u(T\wedge\tau_{\boldsymbol{D}}, \xi_{T\wedge\tau_{\boldsymbol{D}}}) \Big| \xi_s = i \right] - u(s,i) \quad \forall\, (s,i) \in [0,T]\times\boldsymbol{S},\ \boldsymbol{D} \subseteq \boldsymbol{S}.$$

where $c(i,\pi_v) := \int_{\boldsymbol{A}(i)} c(i,a)\pi_v(da|i)$, $q(j|i,\pi_t) := \int_{\boldsymbol{A}(i)} q(j|i,a)\pi_t(da|i)$ for $i,j \in \boldsymbol{S}, t \geq 0$.

Proof. By Condition 3.1(c), we have $2tc(i,a) \le 2t\delta(i) \le 2t\sqrt{\ln V_0(i)} \le t^2 + \ln V_0(i)$ for all $t \ge 0$, thus $e^{2Tc(i,a)} \le e^{T^2} V_0(i)$ for all $i \in \mathbf{S}$, which verifies Condition 2.1(c) with $M_0 := e^{T^2}$. Thus, in the proof of Theorem 2.1, replacing T with $T \wedge \tau_{\mathbf{D}}$, we see that this lemma is also true. □

Lemma 3.3. *If Conditions 3.1 and 3.2 are satisfied, then the following assertions hold.*

(a) $E_i^\pi \left[e^{\int_0^{\tau_{i_0}} \int_{\boldsymbol{A}} c(\xi_v,a)\pi(da|\omega,v)dv} \right] \le E_i^\pi \left[e^{\int_0^{\tau_{i_0}} \delta(\xi_v)dv} \right] \le V_0(i)$ *for* $i \in \boldsymbol{S}$ *and* $\pi \in \Pi$.

(b) $E_i^\pi[\tau_{i_0}] \le \frac{V_0(i)}{\delta_*}$, *and* $P_i^\pi(\tau_{i_0} < \infty) = 1$, *for all* $i \in \boldsymbol{S}$ *and* $\pi \in \Pi$.

Proof. (a) Obviously, the results hold for $i = i_0$. For any $i \ne i_0$, since $V_0 \ge 1$, by Lemma 3.2(a), we have

$$\begin{aligned}
& E_i^\pi \left[e^{\int_0^{T\wedge\tau_{i_0}} \delta(\xi_v)dv} \right] - V_0(i), \\
\le\ & E_i^\pi \left[e^{\int_0^{T\wedge\tau_{i_0}} \delta(\xi_v)dv} V_0(\xi_{T\wedge\tau_{i_0}}) \right] - V_0(i) \\
=\ & E_i^\pi \left[\int_0^{T\wedge\tau_{i_0}} e^{\int_0^t \delta(\xi_v)dv} \left(\delta(\xi_t)V_0(\xi_t) + \sum_{j\in\mathbf{S}} V_0(j) q(j|\xi_t,\pi_t) \right) dt \right] \\
\le\ & b_0 E_i^\pi \left[\int_0^{T\wedge\tau_{i_0}} e^{\int_0^t \delta(\xi_v)dv} I_{\{i_0\}}(\xi_t)dt \right] = 0,
\end{aligned}$$

which, together with letting $T \to \infty$, proves this lemma.

(b) Since $\delta_* = \inf_{i\ne i_0} \delta(i) > 0$, by (a) and the Jensen inequality, we have

$$\delta_* E_i^\pi[\tau_{i_0}] \le e^{\delta_* E_i^\pi[\tau_{i_0}]} \le E_i^\pi \left[e^{\int_0^{\tau_{i_0}} \delta(\xi_v)dv} \right] \le V_0(i),$$

which implies (b). □

Remark 3.2. *The reference state i_0 in Lemma 3.3 directly comes from Condition 3.1(a). However, the reference state i_0 in [74] has been determined by the condition "$V(i_0) \ge 1 + \frac{b}{\delta}$", where the constants δ and b are the same as in Assumption A2 in [74]. Example 3.1 below shows that the condition "$V(i_0) \ge 1 + \frac{b}{\delta}$" is not used to get a reference state.*

Example 3.1. Let $\mathbf{S} := \{0, 1, 2\}$, all $\mathbf{A}(i)$ be singleton sets, and then there is a unique stationary policy, say f. Moreover, let $q(0|0, f(0)) = -1, q(1|0, f(0)) = 1, q(2|0, f(0)) = 0; q(0|1, f(1)) = 8, q(1|1, f(1)) = -9, q(2|1, f(1)) = 1; q(0|2, f(2)) = 0$, $q(1|2, f(2)) = 8$, $q(2|2, f(2)) = -8$.

Obviously, Assumption A2 in [74] are satisfied for the $V(i) := 1 + i$ for $i \in S$, $C := \{0, 1\}$ (or $\{0\}$), $\delta := \frac{4}{3}, b := \frac{11}{3}$. It follows from this example that $1 + \frac{b}{\delta} = 3.75 > V(i)$ for all $i \in \mathbf{S}$, and thus this example does not have any reference state for [74].

However, since Conditions 3.1 and 3.2 above are also satisfied for $V(i) := 1+i$ for $i \in \mathbf{S}$ and $i_0 = 0$, the state "0" is a reference one for this paper.

To characterize some optimality results for the risk-sensitive finite-horizon CTMDPs, we introduce the following condition.

Condition 3.3. (a) *For any fixed $i, j \in \boldsymbol{S}$, $q(j|i, a)$ and $c(i, a)$ is continuous in $a \in \boldsymbol{A}(i)$;*

(b) *For any given $i \in \boldsymbol{S}$, the convergence of $\sum_{j \in \boldsymbol{S}} V_0(j) q(j|i, a)$ holds uniformly in $a \in \boldsymbol{A}(i)$.*

Remark 3.3. *Obviously, Condition 3.3(b) is not required when $\sum_{j \in \boldsymbol{S}(i)} q(j|i, a) = 0$ for $a \in \boldsymbol{A}(i)$,where $\boldsymbol{S}(i)$ is a finite subset of $\boldsymbol{S}$,which may depend on any given $i \in \boldsymbol{S}$. Condition 3.3 implies that $\sum_{j \in \boldsymbol{S}} V_0(j) q(j|i, a)$ is continuous in $a \in \boldsymbol{A}(i)$, and so slightly stronger than the well known continuity-compactness conditions [43, 74, 73]. In fact, it is required not only for the following facts from previous Chapter but also for the continuity of $\varphi'(t, i)$ at everywhere $t \geq 0$ (instead of at almost everywhere $t \geq 0$ in [74]), see Remark 3.5 below for more details.*

Theorem 3.1. *Under Conditions 3.1 and 3.2, and 3.3, for any $T > 0$, the following assertions hold.*

(a) *The value function $\varphi(t, i)$ is the unique solution in $C^1_{V_0, V_1}([0, T] \times \boldsymbol{S})$ of the*

following risk-sensitive finite-horizon optimality equation:

$$\begin{cases} \varphi'(t,i) = \inf_{a\in \boldsymbol{A}(i)}[c(i,a)\varphi(t,i) + \sum_{j\in \boldsymbol{S}} \varphi(t,j)q(j|i,a)], \\ \varphi(0,i) = 1, \end{cases} \tag{3.3}$$

for each $(t,i) \in [0,T] \times \boldsymbol{S}$*, and* $\varphi'(t,i)$ *is continuous in* $t \geq 0$ *(for any fixed* $i \in \boldsymbol{S}$*).*

(b) *There exists a deterministic Markov policy* $\pi^* = f^*(t,i) \in \Pi_m^d$ *such that*

$$\varphi'(t,i) = c(i, f^*(t,i))\varphi(t,i) + \sum_{j\in \boldsymbol{S}} \varphi(t,j)q(j|i,f^*(t,i)) \quad \forall\ (t,i) \in [0,T] \times \boldsymbol{S}$$

and $\varphi(t,i) = \varphi(t,i,\pi^*) = \inf_{\pi\in\Pi_m^d} \varphi(t,i,\pi) = \inf_{\pi\in\Pi_m^r} \varphi(t,i,\pi)$ *for all* $i \in \boldsymbol{S}$ *and* $t \in [0,T]$*.*

(c) *(The multiplicative dynamic programming principle.) For any subset* $\boldsymbol{D}$ *of* $\boldsymbol{S}$*, the value function* $\varphi(i,t)$ *can be represented as*

$$\begin{aligned} \varphi(t,i) &= \inf_{f\in\Pi_m^d} E_i^\pi \left[e^{\int_0^{t\wedge\tau_{\boldsymbol{D}}} c(\xi_v, f(v,\xi_v))dv} \varphi(t - t\wedge\tau_{\boldsymbol{D}}, \xi_{t\wedge\tau_{\boldsymbol{D}}}) \right] \\ &= \inf_{\pi\in\Pi_m^r} E_i^\pi \left[e^{\int_0^{t\wedge\tau_{\boldsymbol{D}}} c(\xi_v, \pi_v)dv} \varphi(t - t\wedge\tau_{\boldsymbol{D}}, \xi_{t\wedge\tau_{\boldsymbol{D}}}) \right] \ \forall\ i \in \boldsymbol{S}, t \geq 0. \end{aligned} \tag{3.4}$$

Proof. (a)-(b): Since Condition 2.1(c) has been verified in the proof of Lemma 3.2, the first part of (a) comes from Theorem 2.2 and an obvious change of time. To show the second part of (a), for any fixed $i \in \mathbf{S}, t_1, t_2 \in [0,T], \varepsilon > 0$, Condition 3.3 together with Lemma 4.1 in [52] ensures the existence of $a(i,t_1,t_2) \in \mathbf{A}(i)$ and a finite subset $\mathbf{S}(i)(\ni i)$ of $\mathbf{S}$ such that $\sup_{a\in\mathbf{A}(i)} \sum_{j\notin\mathbf{S}(i)} |\varphi(t_1,j) - \varphi(t_2,j)||q(j|i,a)| = \sum_{j\notin\mathbf{S}(i)} |\varphi(t_1,j) - \varphi(t_2,j)||q(j|i,a(i,t_1,t_2))|$, and $\sum_{j\notin\mathbf{S}(i)} V_0(j)q(j|i,a) < \varepsilon$ for all

$a \in \mathbf{A}(i)$. Thus, for any $t_1, t_2 \in [0,T]$, by (3.3), we have

$$\begin{aligned}
|\varphi'(t_1,i)-\varphi'(t_2,i)| &\leq \|c(i,\cdot)\||\varphi(t_1,i)-\varphi(t_2,i)| + \sup_{a\in\mathbf{A}(i)}\sum_{j\in\mathbf{S}}|\varphi(t_1,j)-\varphi(t_2,j)||q(j|i,a)| \\
&\leq \|c(i,\cdot)\||\varphi(t_1,i)-\varphi(t_2,i)| + \sum_{j\in\mathbf{S}(i)}|\varphi(t_1,j)-\varphi(t_2,j)|q^*(i) \\
&\quad +2\|\varphi\|_{V_0}\sum_{j\notin\mathbf{S}(i)}V_0(j)q(j|i,a(i,t_1,t_2)) \\
&\leq \|c(i,\cdot)\||\varphi(t_1,i)-\varphi(t_2,i)| + \sum_{j\in\mathbf{S}(i)}|\varphi(t_1,j)-\varphi(t_2,j)|q^*(i) + 2\|\varphi\|_{V_0}\varepsilon
\end{aligned}$$

which,together with the continuity of $\varphi(t,i)$ in t and the finiteness of $\|c(i,\cdot)\|, \mathbf{S}(i)$ and $q^*(i)$, implies the second part of (a). Part (b) is also from Theorem 2.2.

(c) Let $\pi = f(t,i)$ be any Markov policy in Π_m^d, and $\pi^* = f^*(t,i)$ a fixed deterministic Markov policy from part (b). Define a policy $\hat{\pi}$ by

$$\hat{\pi}(da|\omega,s) = I_{\{t\wedge\tau_{\mathbf{D}}>s\}}\delta_{f(s,\xi_s(\omega))}(da) + I_{\{t\wedge\tau_{\mathbf{D}}\leq s\}}\delta_{f^*(s,\xi_s(\omega))}(da).$$

Let $\mathcal{F}_{t\wedge\tau_{\mathbf{D}}}$ be the algebra,which is generated by the stopping time $t\wedge\tau_{\mathbf{D}}$ with respective the filtration $\mathcal{F}_s := \sigma(\xi_v, v\leq s)$. Then, since $\varphi(t,i) = \varphi(t,i,\pi^*)$ for $t\in[0,T]$ and $i\in\mathbf{S}$, we have

$$\begin{aligned}
\varphi(t,i) &\leq E_i^{\hat{\pi}}\left[e^{\int_0^t\int_{\mathbf{A}}c(\xi_s,a)\hat{\pi}(da|\omega,s)ds}\right] \\
&= E_i^{\hat{\pi}}\left[e^{\int_0^{t\wedge\tau_{\mathbf{D}}}c(\xi_s,f(s,\xi_s))ds}E_i^{\hat{\pi}}\left[e^{\int_{t\wedge\tau_{\mathbf{D}}}^t c(\xi_s,f^*(s,\xi_s))ds}|\mathcal{F}_{t\wedge\tau_{\mathbf{D}}}\right]\right] \\
&= E_i^{\pi}\left[e^{\int_0^{t\wedge\tau_{\mathbf{D}}}c(\xi_s,f(s,\xi_s))ds}\varphi(t-t\wedge\tau_{\mathbf{D}},\xi_{t\wedge\tau_{\mathbf{D}}},\pi^*)\right] \\
&= E_i^{\pi}\left[e^{\int_0^{t\wedge\tau_{\mathbf{D}}}c(\xi_s,f(s,\xi_s))ds}\varphi(t-t\wedge\tau_{\mathbf{D}},\xi_{t\wedge\tau_{\mathbf{D}}})\right] \\
&\leq E_i^{\pi}\left[e^{\int_0^{t\wedge\tau_{\mathbf{D}}}c(\xi_s,f(s,\xi_s))ds}\varphi(t-t\wedge\tau_{\mathbf{D}},\xi_{t\wedge\tau_{\mathbf{D}}},\pi)\right] = \varphi(t,i,\pi).
\end{aligned}$$

Taking the infimum over $\pi\in\Pi_m^d$ on the both sides of the above inequality and using (b) again, we see that part (c) holds. □

Theorem 3.2. *Under Conditions 3.1, 3.2 and 3.3, the following assertions hold.*

(a) $\frac{\varphi(t,i)}{\varphi(t,i_0)} \leq V_0(i)$ *for all and* $i\in\boldsymbol{S}$ *and* $t\geq 0$*;*

(b) $\varphi'(t,i) \geq 0$ *and* $\sup_{t\geq 0} \frac{\varphi'(t,i)}{\varphi(t,i_0)} \leq L(i)$ *for all* $i \in \boldsymbol{S}$ *and* $t \geq 0$, *where* $L(i) := 2V_0(i)q^*(i) + b_0$.

Proof. (a) Since $1 = \varphi(0,i) \leq \varphi(t,i)$ for all $i \in \mathbf{S}$ and $t \geq 0$, by Theorem 3.1(c) we have, for each $f \in \Pi_m^d$,

$$\begin{aligned}
\varphi(t,i) &\leq E_i^\pi \left[e^{\int_0^{t\wedge\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \varphi(t - t\wedge\tau_{i_0}, \xi_{t\wedge\tau_{i_0}}) \right] \\
&= E_i^\pi \left[e^{\int_0^{t\wedge\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \varphi(t - t\wedge\tau_{i_0}, \xi_{t\wedge\tau_{i_0}}) I_{\{t\leq\tau_{i_0}\}} \right] \\
&\quad + E_i^\pi \left[e^{\int_0^{t\wedge\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \varphi(t - t\wedge\tau_{i_0}, \xi_{t\wedge\tau_{i_0}}) I_{\{t>\tau_{i_0}\}} \right] \\
&= E_i^\pi \left[e^{\int_0^{t\wedge\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} I_{\{t\leq\tau_{i_0}\}} \right] \\
&\quad + E_i^\pi \left[e^{\int_0^{t\wedge\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \varphi(t - \tau_{i_0}, \xi_{\tau_{i_0}}) I_{\{t>\tau_{i_0}\}} \right] \\
&\leq E_i^\pi \left[e^{\int_0^{t\wedge\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \varphi(t, \xi_{\tau_{i_0}}) I_{\{t\leq\tau_{i_0}\}} \right] \\
&\quad + E_i^\pi \left[e^{\int_0^{t\wedge\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \varphi(t, \xi_{\tau_{i_0}}) I_{\{t>\tau_{i_0}\}} \right] \\
&\leq E_i^\pi \left[e^{\int_0^{\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \varphi(t, \xi_{\tau_{i_0}}) \right] \qquad (3.5) \\
&= \varphi(t,i_0) E_i^\pi \left[e^{\int_0^{\tau_{i_0}} c(\xi_v, f(v,\xi_v))dv} \right],
\end{aligned}$$

which, together with Lemma 3.3(a), completes the proof of part (a).

(b) Since $\varphi(t,i)$ is increasing in t for each $i \in \mathbf{S}$, and thus $\varphi'(t,i) \geq 0$. Moreover,for each $t \geq 0$, by (a) and Theorem 3.1(b) we have

$$\begin{aligned}
0 \leq \frac{\varphi'(t,i)}{\varphi(t,i_0)} &= \inf_{a\in\mathbf{A}(i)} [c(i,a)\frac{\varphi(t,i)}{\varphi(t,i_0)} + \sum_{j\in\mathbf{S}} \frac{\varphi(t,j)}{\varphi(t,i_0)} q(j|i,a)] \\
&\leq \inf_{a\in\mathbf{A}(i)} [\delta(i)V_0(i) + \sum_{j\in\mathbf{S}} V_0(j)|q(j|i,a)|] \\
&= \inf_{a\in\mathbf{A}(i)} [\delta(i)V_0(i) + \sum_{j\in\mathbf{S}} V_0(j)q(j|i,a) - 2q(i|i,a)V_0(i)] \\
&\leq 2q^*(i)V_0(i) + b_0.
\end{aligned}$$

and so part (b) follows. □

Remark 3.4. *If the $I_{\{i_0\}}$ in Condition 3.1(i) is replaced with the indicator function $I_{\boldsymbol{C}}$ of some finite subset $\boldsymbol{C}$ of $\boldsymbol{S}$, as the proof of (3.5), we can prove that*

$$E_i^\pi\left[e^{\int_0^{\tau_{\boldsymbol{C}}}\delta(\xi_v)dv}\right] \le V_0(i), \text{ and } \varphi(t,i) \le E_i^\pi\left[e^{\int_0^{\tau_{\boldsymbol{C}}}c(\xi_v,\pi_v)dv}\right]\varphi(t,i_0(t)) \quad \forall\ t \ge 0 \tag{3.6}$$

where the states $i_0(t)$ (depending on $t \ge 0$) are determined by $\varphi(t,i_0(t)) := \max_{i\in\boldsymbol{C}}\varphi(t,i)$. However, we can not prove Lemma 3.3 and Theorem 3.2 by (3.6). In fact, from (3.6) we cannot establish the existence of some fixed $i_0 \in \boldsymbol{C}$ such that :

$$E_i^\pi\left[e^{\int_0^{\tau_{i_0}}\delta(\xi_v)dv}\right] \le V_0(i),\ \frac{\varphi(t,i)}{\varphi(t,i_0)} \le E_i^\pi\left[e^{\int_0^{\tau_{i_0}}c(\xi_v,\pi_v)dv}\right]\ \forall\ t \ge 0,$$

which are required in [74] and our arguments below. This is because $\tau_{i_0} \ge \tau_{\boldsymbol{C}}$ and the states $i_0(t)$ may change with $t \ge 0$.

3.2 On the risk-sensitive average optimality equation

In this section, we will establish the existence of a solution to the RS-AOE for the risk-sensitive average CTMDPs without loss of generalization. We suppose that $\mathbf{S} = \{0, 1, \cdots\}$ (the set of all nonnegative integers).

To begin with, we need some notation given as follows: For each $n \ge 1$, let

$$c_n(i,a) := \begin{cases} c(i,a), & \text{for } i \in \{0,\dots,n\}, a \in \mathbf{A}(i), \\ 0, & \text{otherwise}, \end{cases} \tag{3.7}$$

Using Theorems 3.1 and 3.2, we give the extensions of the corresponding ones in [74] to the unbounded transition and cost rates.

Theorem 3.3. Under Conditions 3.1, 3.2 and 3.3, for each $n \ge i_0$, the followings hold.

(a) There exists a solution (ρ_n, ψ_n) in $[0, L(i_0)] \times B_{V_0}(\mathbf{S})$ to the following equation

$$\rho_n \psi_n(i) = \inf_{a \in \mathbf{A}(i)} \{c_n(i,a)\psi_n(i) + \sum_{j \in \mathbf{S}} \psi_n(j) q(j|i,a)\}, \psi_n(i_0) = 1, \quad i \in \mathbf{S}. \tag{3.8}$$

(b) $\rho_n \leq \inf_{\pi \in \Pi} J(i, \pi)$ for all $i \in \mathbf{S}$.

(c) There is a policy $f_n^* \in F$ such that

$$\psi_n(i) = E_i^{f_n^*} \left[e^{\int_0^{\tau_{i_0}} (c_n(\xi_t, f_n^*(\xi_t)) - \rho_n) dt} \right] = \inf_{f \in \Pi_m^d} E_i^f \left[e^{\int_0^{\tau_{i_0}} (c_n(\xi_t, f(t,\xi_t)) - \rho_n) dt} \right] \quad \forall i \in \mathbf{S}.$$

Proof. (a) For any fixed $n \geq 1$, let

$$\varphi_n(t,i) := \inf_{f \in \Pi_m^d} E_i^f \left[e^{\int_0^t \int_{\mathbf{A}} c_n(\xi_v, f(v,\xi_v)) dv} \right], \text{ and } \hat{\varphi}_n(t,i) := \frac{\varphi_n(t,i)}{\varphi_n(t,i_0)}, \quad \text{for } t \geq 0, \ i \in \mathbf{S} \tag{3.9}$$

Then, since $0 \leq c_n \leq c$, the conditions of Theorems 3.1 and 3.2 still hold when c is replaced with c_n. Thus, using Theorem 3.2(a), we have

$$0 \leq \hat{\varphi}_n(t,i) \leq V_0(i) \quad \forall \ t \geq 0.$$

Thus, for any $m \geq 1$, the mean value theorem together with Theorems 3.1(a) gives the existence of $k_0(i,m) \in [m, 2m]$ (depending on the given m and i) such that

$$\hat{\varphi}_n'(k_0(m,i), i) = \frac{\hat{\varphi}_n(2m,i) - \hat{\varphi}_n(m,i)}{2m - m} \to 0 \ \ (\text{as } m \to \infty). \tag{3.10}$$

Since $\mathbf{S}$ is denumerable, the diagonalization arguments as well as (3.10) ensures the existence of a subsequence of $\{k_1\}$ of the $\{k_0(i,m), m \geq 1, i \in \mathbf{S}\}$ such that

$$\lim_{k_1 \to \infty} \hat{\varphi}_n'(k_1, i) = 0, \ \hat{\varphi}_n(k_1, i_0) \equiv 1, \quad \text{for all } i \in \mathbf{S}.$$

which, together with the boundedness of $|\hat{\varphi}_n(k_1, i)|$ in k_1 and the denumerability

of $i \in \mathbf{S}$, gives the existence of a subsequence $\{k_2\}$ of the $\{k_1\}$ such that the limits

$$\lim_{k_2\to\infty} \hat{\varphi}_n(k_2, i) =: \psi_n(i) \in [0, V_0(i)], \tag{3.11}$$

exist for all $i \in \mathbf{S}$, and $\psi_n(i_0) = 1$.

Furthermore, Theorem 3.2(b) together with (3.11) guarantees the existence of a subsequence $\{k_3\}$ of the $\{k_2\}$ ensuring the existence of the following limit

$$\lim_{k_3\to\infty} \frac{\varphi_n'(k_3, i_0)}{\varphi_n(k_3, i_0)} =: \rho_n \in [0, L(i_0)]. \tag{3.12}$$

Hence, along the sequence $\{k_3\}(\subset \{k_2\} \subset \{k_1\} \subset \{k_0(i, m), m \geq 1, i \in \mathbf{S}\})$, we have

$$\lim_{k_3\to\infty} \left[\hat{\varphi}_n'(k_3, i) + \hat{\varphi}_n(k_3, i)\frac{\varphi_n'(k_3, i_0)}{\varphi_n(k_3, i_0)}\right] = \rho_n\psi_n(i) \qquad \forall\ i \in \mathbf{S}. \tag{3.13}$$

On the other hand, for each $k_3 \in \{k_3\}$, using the definition of $\hat{\varphi}_n(t, i)$ in (3.9) and Theorem 3.1(b) with (c, φ) replaced with the corresponding (c_n, φ_n), a direct calculation gives for each $(i, a) \in \mathbb{K}$

$$\begin{aligned} \hat{\varphi}_n'(k_3, i) + \hat{\varphi}_n(k_3, i)\frac{\varphi_n'(k_3, i_0)}{\varphi_n(k_3, i_0)} &= \inf_{a\in\mathbf{A}(i)}\{c_n(i, a)\hat{\varphi}_n(k_3, i) + \sum_{j\in\mathbf{S}} \hat{\varphi}_n(k_3, j)q(j|i, a)\} \\ &\leq c_n(i, a)\hat{\varphi}_n(k_3, i) + \sum_{j\in\mathbf{S}} \hat{\varphi}_n(k_3, j)q(j|i, a) \end{aligned} \tag{3.14}$$

which, together with the dominated convergence theorem and (3.11)-(3.13), implies

$$\rho_n\psi_n(i) \leq \inf_{a\in\mathbf{A}(i)}\{c_n(i, a)\psi_n(i) + \sum_{j\in\mathbf{S}} \psi_n(j)q(j|i, a)\} \qquad \forall\ i \in \mathbf{S}. \tag{3.15}$$

Moreover, for each given k_3 and $i \in \mathbf{S}$, by Condition 3.3 and (3.13)(3.14), there

exists $a(i,k_3) \in \mathbf{A}(i)$ (depending on $k_3, i \in \mathbf{S}$) such that

$$\hat{\varphi}_n'(k_3,i) + \hat{\varphi}_n(k_3,i)\frac{\varphi_n'(k_3,i_0)}{\varphi_n(k_3,i_0)} = c_n(i,a(i,k_3))\hat{\varphi}_n(k_3,i) + \sum_{j\in\mathbf{S}} \hat{\varphi}_n(k_3,j)q(j|i,a(i,k_3)) \quad (3.16)$$

Since $\mathbf{A}(i)$ is compact, there exist a subsequence of $\{k_4\}$ of of $\{k_3\}$ and $a'(i) \in \mathbf{A}(i)$ such that $\lim_{k_4\to\infty} a(i,k_4) = a'(i)$. Thus, replacing k_3 in (3.16) with k_4 and then letting $k_4 \to \infty$, by (3.13) and (3.16) we have

$$\begin{aligned} \rho_n\psi_n(i) &= c_n(i,a'(i))\psi_n(i) + \sum_{j\in\mathbf{S}} \psi_n(j)q(j|i,a'(i)) \\ &\geq \inf_{a\in\mathbf{A}(i)} \{c_n(i,a)\psi_n(i) + \sum_{j\in\mathbf{S}} \psi_n(j)q(j|i,a)\}, \end{aligned} \quad (3.17)$$

which, together with (3.14) and (3.11)-(3.12), gives (a).

(b) Fix any $i \geq n+1$ with $n \geq i_0$, and let $\tau(n) := \tau_{\{0,1,\ldots,i_0,\ldots,n\}}$ be the hitting time. Then, by Lemma 3.3(b) and $\tau(n) \leq \tau_{i_0}$, $P_i^\pi(\tau(n) < \infty) = 1$. Moreover, since $\varphi_n(0,i) \equiv 1$ and $\varphi_n(t,i)$ is increasing in $t \geq 0$, using (3.7)and Theorem 3.1(c) (with c replacing by c_n here) as well as the fact that $c_n(\xi_v,\pi_v) \equiv 0$ for $v < \tau(n)$ and $\pi \in \Pi_m^r$, we have, for each $\pi \in \Pi_m^r$,

$$\begin{aligned} \varphi_n(t,i) &\leq E_i^\pi \left[e^{\int_0^{t\wedge\tau(n)} c_n(\xi_v,\pi_v)dv} \varphi_n(t - t\wedge\tau(n), \xi_{t\wedge\tau(n)}) \right] \\ &= E_i^\pi \left[e^{\int_0^{t\wedge\tau(n)} c_n(\xi_v,\pi_v)dv} \varphi_n(t - t\wedge\tau(n), \xi_{t\wedge\tau(n)}) I_{\{t\leq\tau(n)\}} \right] \\ &\quad + E_i^\pi \left[e^{\int_0^{t\wedge\tau(n)} c_n(\xi_v,\pi_v)dv} \varphi_n(t - t\wedge\tau(n), \xi_{t\wedge\tau(n)}) I_{\{t>\tau(n)\}} \right] \\ &= E_i^\pi \left[\varphi_n(0,\xi_t) I_{\{t\leq\tau(n)\}} \right] + E_i^\pi \left[\varphi_n(t-\tau(n),\xi_{\tau(n)}) I_{\{t>\tau(n)\}} \right] \\ &\leq 1 + E_i^\pi \left[\varphi_n(t,\xi_{\tau(n)}) I_{\{t>\tau(n)\}} \right] \\ &\leq 1 + \max\{\varphi_n(t,k), k = 0,1,\ldots,n\}, \end{aligned}$$

which,together with Theorem 3.2(a) and $\varphi_n \geq 1$, implies

$$\hat{\varphi}_n(t,i) \leq 1 + \max\{V_0(k), k = 0,1,\ldots,n\} =: K(n) \quad \forall\ t \geq 0, i \in \mathbf{S}.$$

Hence, using (3.11) and Theorem 3.2(a) again, we have

$$\psi_n(i) \le K(n) \quad \forall\, t \ge 0, i \in \mathbf{S}. \tag{3.18}$$

For each $\pi \in \Pi$ and $i \in \mathbf{S}$, using Lemma 3.2(a) and (3.8), we have

$$E_i^\pi \left[e^{\int_0^T \int_{\mathbf{A}} [c_n(\xi_t,a)-\rho_n]\pi(da|\omega,t)dt} \psi_n(\xi_T) \right] - \psi_n(i) \ge 0, \quad \text{for } T > 0.$$

Hence, by (3.18) we have

$$K(n) e^{-T\rho_n} E_i^\pi \left[e^{\int_0^T \int_{\mathbf{A}} c_n(\xi_t,a)\pi(da|\omega,t)dt} \right] \ge \psi_n(i).$$

Taking logarithm in the above sides, dividing by T and letting $T \to \infty$, we obtain

$$\rho_n \le \limsup_{T\to\infty} \frac{1}{T} \ln E_i^\pi \left[e^{\int_0^T \int_{\mathbf{A}} c_n(\xi_t,a)\pi(da|\omega,t)dt} \right] \le \limsup_{T\to\infty} \frac{1}{T} \ln E_i^\pi \left[e^{\int_0^T \int_{\mathbf{A}} c(\xi_t,a)\pi(da|\omega,t)dt} \right],$$

which, together with the arbitrariness of π, completes the proof of (b).

(c) For each $f \in \Pi_m^d$ and $i \in \mathbf{S}$, using Lemma 3.2(b) and (3.8) again, we have

$$E_i^f \left[e^{\int_0^{T\wedge\tau_{i_0}} (c_n(\xi_t,f(t,\xi_t))-\rho_n)dt} \psi_n(\xi_{T\wedge\tau_{i_0}}) \right] - \psi_n(i) \ge 0, \quad \text{for } T > 0.$$

Since ψ_n is bounded and $c_n(i,a) \le c(i,a) \le \delta(i)$ for all $(i,a) \in \mathbb{K}$, by the dominated convergence theorem and Lemma 3.3(a), letting $T \to \infty$, we obtain

$$\begin{aligned} \psi_n(i) &\le E_i^f \left[e^{\int_0^{\tau_{i_0}} [c_n(\xi_t,f(t,\xi_t))-\rho_n]dt} \psi_n(\xi_{\tau_{i_0}}) \right] \\ &= E_i^f \left[e^{\int_0^{\tau_{i_0}} [c_n(\xi_t,f(t,\xi_t))-\rho_n]dt} \right] \\ &\le V_0(i) \end{aligned} \tag{3.19}$$

which, together with the arbitrariness of $\pi \in \Pi_m^d$, implies

$$\psi_n(i) \le \inf_{f\in\Pi_m^d} E_i^f \left[e^{\int_0^{\tau_{i_0}} [c_n(\xi_t,f(t,\xi_t))-\rho_n]dt} \psi_n(\xi_{\tau_{i_0}}) \right] \le V_0(i). \tag{3.20}$$

On the other hand, take $f_n^* \in F$ be a minimizing stationary policy in (3.8), that is,

$$\rho_n \psi_n(i) = c(i, f_n^*(i))\psi_n(i) + \sum_{j \in \mathbf{S}} \psi_n(j) q(j|i, f_n^*(i)) \quad \forall\ i \in \mathbf{S}. \tag{3.21}$$

Using Lemma 3.2(b) and (3.21), as the proof of (3.20) (by the Fatou's lemma again) we have

$$\psi_n(i) \geq E_i^{f_n^*} \left[e^{\int_0^{\tau_{i_0}} (c_n(\xi_t, f_n^*(\xi_t)) - \rho_n) dt} \right].$$

This inequality as well as (3.20) gives (c). □

Remark 3.5. In order to establish the existence of $k_0(m, i)$ for (3.10) to hold, the existence of $\varphi'(t, i)$ of the function $\varphi(t, i)$ at each $t \geq 0$ is required. Otherwise, such $k_0(m, i)$ can not be guaranteed. For example, for some given $i \in \mathbf{S}$ and $n \geq 1$, suppose that

$$\hat{\varphi}_n(t, i) := \begin{cases} 2t - 2, & t \in [1, 2), \\ 4 - 2t, & t \in [2, 4]. \end{cases} \tag{3.22}$$

Take $m = 1$. Then, $\frac{\hat{\varphi}_n(2,i) - \hat{\varphi}_n(1,i)}{2-1} = 0 \neq \hat{\varphi}_n'(k_0(1, i), i)$ for any $k_0(1, i) \in [1, 2)$.

Theorem 3.4. Under Conditions 3.1, 3.2 and 3.3, the followings hold.

(a) There exists a solution (ρ^*, ψ^*) in $[0, L(i_0)] \times B_{V_0}(\mathbf{S})$ to the following *RS-AOE*

$$\psi^*(i_0) = 1, \quad \rho^* \psi^*(i) = \inf_{a \in \mathbf{A}(i)} \{c(i, a)\psi^*(i) + \sum_{j \in \mathbf{S}} \psi^*(j) q(j|i, a)\} \quad \forall\ i \in \mathbf{S}. \tag{3.23}$$

(b) $\rho^* \leq \inf_{\pi \in \Pi} J(i, \pi)$ for all $i \in \mathbf{S}$.

(c) There exists some policy $f^* \in F$ such that

$$(c_1) \qquad \rho^* \psi^*(i) = c(i, f^*(i))\psi^*(i) + \textstyle\sum_{j \in \mathbf{S}} \psi^*(j) q(j|i, f^*(i)) \quad \forall\ i \in \mathbf{S};$$

$$(c_2) \quad \psi^*(i) = \inf_{f\in\Pi_m^d} E_i^f\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f(t,\xi_t))-\rho^*)dt}\right] = E_i^{f^*}\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f^*(\xi_t))-\rho^*)dt}\right]$$

Proof. (a)-(b): Take (ρ_n, ψ_n) as in Theorem 3.3. Then, since $\rho_n \in [0, L(i_0)]$ for all $n \geq i_0$, there exist a subsequence $\{n_1\}$ of $\{n, n \geq i_0\}$ and a constant $\rho^* \in [0, L(i_0)]$ such that $\rho^* = \lim_{n_1\to\infty} \rho_{n_1}$, and $\rho^* \leq \inf_{\pi\in\Pi} J(i,\pi)$ (by Theorem 3.3(b)). Furthermore, since $\psi_n \in B_{V_0}(\mathbf{S})$ and $\psi_n(i_0) = 1$ for all $n \geq i_0$ and $\mathbf{S}$ is denumerable, the diagonalization argument ensures the existence of a subsequence $\{n_2\}$ of $\{n_1\}$ and a function $\psi^* \in B_{V_0}(\mathbf{S})$ satisfying

$$\rho^* = \lim_{n_2\to\infty} \rho_{n_2}, \psi^*(i) = \lim_{n_2\to\infty} \psi_{n_2}(i) \quad \forall i \in S,\ \psi^*(i_0) = 1.$$

Then, replacing n in (3.15) and (3.17) with n_2 and letting $n_2 \to \infty$, get

$$\rho^*\psi^*(i) = \inf_{a\in A(i)} \{c(i,a)\psi^*(i) + \sum_{j\in S} \psi^*(j)q(j|i,a)\} \qquad \forall i \in \mathbf{S}.$$

Thus, we completes the proof of (a) and (b).

(c). For any given $f \in \Pi_m^d$, since $c_{n_2} \leq c$ and $\rho_{n_2} \geq 0$, Theorem 3.3(c) gives that

$$\psi_{n_2}(i) \leq E_i^f\left[e^{\int_0^{\tau_{i_0}}(c_{n_2}(\xi_t,f(t,\xi_t))-\rho_{n_2})dt}\right] \leq E_i^f\left[e^{\int_0^{\tau_{i_0}} c(\xi_t,f(t,\xi_t))dt}\right] \leq V_0(i) \quad \forall\ n_2 \geq 1.$$

Thus, using the dominated convergence theorem and letting $n_2 \to \infty$, we have

$$\psi^*(i) \leq E_i^f\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f(t,\xi_t))-\rho^*)dt}\right],$$

and hence,

$$\psi^*(i) \leq \inf_{f\in\Pi_m^d} E_i^\pi\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f(t,\xi_t))-\rho^*)dt}\right] \quad \forall\ i \in \mathbf{S}. \tag{3.24}$$

Taking $f_{n_2}^*$ as in (3.21). Since $f_{n_2}^*(i) \in \mathbf{A}(i)$ belongs to the compact $\mathbf{A}(i)$ (for each $i \in \mathbf{S}$) and $\mathbf{S}$ is denumerable, there exists a subsequence $\{n_3\}$ of $\{n_2\}$ and

$f^* \in F \subset \Pi_m^d$ such that

$$c(i, f^*(i)) = \lim_{n_3 \to \infty} c(i, f^*_{n_3}(i)), q(j|i, f^*(i)) = \lim_{n_3 \to \infty} q(j|i, f^*_{n_3}(i)) \ \ \forall\ i, j \in \mathbf{S}.$$

Replacing n in (3.21) with n_3 and then letting $n_3 \to \infty$, we have

$$\rho^* \psi^*(i) = c(i, f^*(i))\psi^*(i) + \sum_{j \in \mathbf{S}} \psi^*(j) q(j|i, f^*(i)) \quad \forall\ i \in \mathbf{S},$$

which, implies (c_1) and also (by the extended Feymanm-Kac's formula)

$$\psi^*(i) = E_i^{f^*} \left[e^{\int_0^{T \wedge \tau_{i_0}} (c(\xi_t, f^*(\xi_t)) - \rho^*) dt} \psi^*(\xi_{T \wedge \tau_{i_0}}) \right], \quad \text{for } T > 0. \tag{3.25}$$

Hence, using Fatou's lemma and noting $\psi^*(x_{\tau_{i_0}}) = \psi^*(i_0) = 1$, letting $T \to \infty$ we have

$$\psi^*(i) \geq E_i^{f^*} \left[e^{\int_0^{\tau_{i_0}} [c(\xi_t, f^*(\xi_t)) - \rho^*] dt)} \right] \geq \inf_{f \in \Pi_m^d} E_i^{\pi} \left[e^{\int_0^{\tau_{i_0}} (c(\xi_t, f(t, \xi_t)) - \rho^*) dt} \right], \tag{3.26}$$

which, together with (3.24) gives (c_2). □

3.3 Existence of risk-sensitive average optimal policies

In this section, we will prove the existence of a risk-sensitive average optimal stationary policy using the RS-AOE. To do so, besides Conditions 3.1, 3.2 and 3.3, which are assumed to hold throughout this section, we need the following condition, whose necessity will be illustrated in the example.

Condition 3.4. $\inf_{i \in S} \psi^*(i) > 0$, where ψ^* is from Theorem 3.4.

Condition 3.4 is new. We will show that, under Condition 3.4 and the conditions in Theorem 3.4, a risk-sensitive average optimal stationary policy exists. Before proving this, we provide some sufficient conditions for the verification of Condition 3.4.

Proposition 3.1. Each one of the following conditions $(a)-(c)$ implies Condition 3.4.

(a) The state space $\mathbf{S}$ is finite.

(b) Suppose that $\inf_{a\in\mathbf{A}(i)} q(i,a) > 0$ for all $i \neq i_0$, and there exist a nonnegative bounded function V_2 on $\mathbf{S}$ satisfying

$$\sum_{j\neq i_0} q(j|i,a)V_2(j) \leq -1 \quad \forall\ a \in \mathbf{A}(i), i \neq i_0,\ V_2(i_0) := 0.$$

(c) (*Stochastic monotonicity condition*). For any $f \in F$, $c(i, f(i))$ is increasing in $i \in \mathbf{S}$; and $\sum_{j\geq k} q(j|i, f(i)) \leq \sum_{j\geq k} q(j|i+1, f(i+1))$ for all $i, k \in \mathbf{S}$ with $k \neq i+1$; and $i_0 = 0$.

Proof. (a) Take f^* and ρ^* as in Theorem 3.4. $\psi^*(i) = E_i^{f^*}\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f^*(\xi_t))-\rho^*)dt}\right] > 0$ for all $i \in \mathbf{S}$. Thus, it is obvious that (a) implies Condition 3.4.

(b) Let $M := \sup_{i\in\mathbf{S}} V_2(i) < \infty$ (by the condition). Then, as the proof of Lemma 6.1.5 in [3], we have $E_i^{f^*}(\tau_{i_0}) \leq V_2(i) \leq M$. Therefore, by the Jensen inequality and $c \geq 0$, we have

$$\psi^*(i) = E_i^{f^*}\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f^*(\xi_t))-\rho^*)dt}\right] \geq e^{-\rho^* E_i^{f^*}(\tau_{i_0})} \geq e^{-\rho^* M} \geq e^{-L(i_0)M} \quad \forall\ i \in \mathbf{S},$$

which also verifies Condition 3.4

(c) Obviously, it suffices to show that ψ^* is increasing on $\mathbf{S}$. First, for any nonnegative increasing function u on $\mathbf{S}$ and $t \geq 0$, by Theorem 7.3.4 and Proposition 7.3.2 in [3], we see that $\sum_{j\geq k} P_i^{f^*}(\xi_t = j)$ is nondecreasing in $i \in \mathbf{S}$ (for every fixed $k \in \mathbf{S}$ and $t \geq 0$), which together with Proposition 7.3.1 in [3], implies that $E_i^{f^*}(u(\xi_t))$ is increasing in $i \in \mathbf{S}$. For any given $n \geq 1$ and $T > 0$, let $\hat{\xi}_k := \xi_{\frac{k}{2^n}T}, k = 0, 1, \ldots, 2^n$. Then, since $\{\xi_t, t \geq 0\}$ is right continuous, we have (by the dominated convergence theorem)

$$\psi^*(i) = E_i^{f^*}\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f^*(\xi_t))-\rho^*)dt}\right] = \lim_{T\to\infty}\lim_{n\to\infty} E_i^{f^*}\left[e^{\sum_{k=0}^{2^n} I_{\{\hat{\xi}_k\neq 0\}}[c(\hat{\xi}_k,f^*(\hat{\xi}_k))-\rho^*]\frac{T}{2^n}}\right].$$

Thus, the rest needs to show that $g(i) := E_i^{f^*}\left[e^{\sum_{k=0}^{2^n} I_{\{\hat{\xi}_k \neq 0\}}[c(\hat{\xi}_k, f^*(\hat{\xi}_k))-\rho^*]\frac{T}{2^n}}\right]$ is increasing in $i \in \mathbf{S}$. Fix any $n \geq 1$, and let $\bar{u}(j) := e^{I_{\mathbf{S}\setminus\{0\}}(j)(c(j,f^*(j))-\rho^*)\frac{T}{2^n}}$ for all $j \neq 0$ and $\bar{u}(0) := 0$. Thus, the $\bar{u}(j)$ is increasing in $j \in \mathbf{S}$, and so is $E_i^{f^*}[\bar{u}(\hat{\xi}_{2^n})]$ in $i \in \mathbf{S}$. Moreover,

$$g(i) = E_i^{f^*}\left[\Pi_{k=0}^{2^n}\bar{u}(\hat{\xi}_k)\right] = E_i^{f^*}\left[\Pi_{k=0}^{2^n-1}\bar{u}(\hat{\xi}_k)E^{f^*}_{\hat{\xi}_{2^n-1}}\left[\bar{u}(\hat{\xi}_{2^n})\right]\right].$$

Let $G^1(j) := \bar{u}(j)E_j^{f^*}[\bar{u}(\hat{\xi}_{2^n})]$ for $j \in \mathbf{S}$. Then, G^1 is nonnegative and increasing on $\mathbf{S}$, and

$$g(i) = E_i^{f^*}\left[\Pi_{k=0}^{2^n-2}\bar{u}(\hat{\xi}_k)G^1(\hat{\xi}_{2^n-1})\right].$$

Let $G^{m+1}(j) := \bar{u}(j)G^m(j)$ for $j \in \mathbf{S}$ and $m = 1, \cdots, 2^n - 1$. Then, by induction we see that all G^{m+1} are nonnegative and increasing on $\mathbf{S}$, and so is $g(i) = E_i^{f^*}[G^{2^n}(\hat{\xi}_1)]$ in $i \in \mathbf{S}$. □

We now present our main result as follows.

Theorem 3.5. Under Conditions 3.1, 3.2, 3.3, and 3.4, the followings hold.

(a) There exists a solution (ρ^*, ψ^*) in $[0, L(i_0)] \times B^+_{V_0}(\mathbf{S})$ to the *RS-AOE*:

$$\psi^*(i_0) = 1, \quad \rho^*\psi^*(i) = \inf_{a\in\mathbf{A}(i)}\{c(i,a)\psi^*(i) + \sum_{j\in\mathbf{S}}\psi^*(j)q(j|i,a)\} \quad \forall i \in \mathbf{S}, \tag{3.27}$$

where $B^+_{V_0}(\mathbf{S}) := \{\psi \in B_{V_0}(\mathbf{S}) : \inf_{i\in\mathbf{S}}\psi(i) > 0\}$.

(b) $\rho^* = \inf_{\pi\in\Pi} J(i,\pi)$ for all $i \in \mathbf{S}$.

(c) There exists some policy $f^* \in F$ such that

(c_1) $\rho^*\psi^*(i) = c(i, f^*(i))\psi^*(i) + \sum_{j\in\mathbf{S}}\psi^*(j)q(j|i, f^*(i)) \quad \forall\ i \in \mathbf{S}$;

(c_2) $\psi^*(i) = E_i^{f^*}\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t, f^*(\xi_t))-\rho^*)dt}\right] = \inf_{f\in\Pi_m^d} E_i^f\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t, f(t,\xi_t))-\rho^*)dt}\right]$;

(c_3) $J(i, f^*) = \rho^* = \inf_{\pi\in\Pi} J(i, \pi)$ for all $i \in \mathbf{S}$, which means that f^* is optimal.

Proof. Using the notation and results in Theorem 3.4, we only need to show that $\rho^* \geq J(i, f^*)$ for all $i \in \mathbf{S}$. Indeed, for each $i \in \mathbf{S}$, by Lemma 3.2 and Assumption 3.4 that $\underline{\psi}^* := \inf_{i\in\mathbf{S}} \psi^*(i) > 0$, we have

$$\psi^*(i) = E_i^{f^*}\left[e^{\int_0^T (c(\xi_t, f^*(\xi_t)) - \rho^*)dt}\psi^*(\xi_T)\right] \geq \underline{\psi}^* E_i^{f^*}\left[e^{\int_0^T (c(\xi_t, f^*(\xi_t)) - \rho^*)dt}\right] \quad \forall\, T > 0.$$

Therefore,

$$\ln \psi^*(i) \geq \ln \underline{\psi}^* + \ln E_i^{f^*}\left[e^{\int_0^T c(\xi_t, f^*(\xi_t))dt}\right] - T\rho^*, \quad \text{for } T > 0.$$

which, implies that

$$\rho^* \geq \limsup_{T\to\infty} \frac{1}{T} \ln E_i^{f^*}\left[e^{\int_0^T c(\xi_t, f^*(\xi_t))dt}\right] = J(i, f^*).$$

□

3.4 A policy iteration algorithm and finite-approximation

We have shown the existence of an optimal stationary policy above. In this section, we focus on the computational approach for finding optimal stationary policies.

Under Conditions 3.1-3.3 and 3.4, for each $f \in F$, by taking $\mathbf{A}(i) := \{f(i)\}$ for all $i \in \mathbf{S}$, it follows from Theorem 3.5 that $J(i, f)$ is independent of states i (i.e., a constant denoted by ρ^f), which together with the function $\psi^f(i) := E_i^f\left[e^{\int_0^{\tau_{i_0}} (c(\xi_t, f(\xi_t)) - \rho^f)dt}\right] \leq V_0(i) (i \in \mathbf{S})$, solves the following multiplicative Poisson

equation

$$\rho\psi(i) = c(i, f(i))\psi(i) + \sum_{j\in\mathbf{S}} \psi(j)q(j|i, f(i)) \quad \forall i \in \mathbf{S}, \text{ with } \psi(i_0) = 1. \quad (3.28)$$

To establish the uniqueness of a solution to the Poisson equation (3.28) and the RS-AOE (3.27), we introduce the following condition.

Condition 3.5. $\sup_{i\in\mathbf{S}} \psi^f(i) < \infty$ for each $f \in F$, $\psi^f(i) := E_i^f \left[e^{\int_0^{\tau_{i_0}} (c(\xi_t, f(\xi_t)) - \rho^f)dt}\right]$.

Remark 3.6. Since $\psi^f(i) \geq \psi^*$ for any $f \in F$, Conditions 3.4 and 3.5 together with Theorem 3.5(c) implies that the ψ^* in the solution (ρ^*, ψ^*) to the RS-AOE needs to be a bounded, positive function which is uniformly bounded away from zero. As mentioned in Remark 5.3 in [43], it is unsolved to show the existence of such a solution. Obviously, Conditions 3.4 and 3.5 are satisfied when $\mathbf{S}$ is finite. For the case of infinitely denumerable states we next give suitable conditions and examples for the verifications of Conditions 3.4 and 3.5.

Proposition 3.2. Suppose that $q_* := \inf_{a\in\mathbf{A}(i), i\neq i_0} q(i, a) > 0$, and there exist a nonnegative bounded function V_3 on $\mathbf{S}$ and a constant $\hat{\delta} > 0$ such that

$$\hat{\delta} < q_*, \ V_3(i_0) = 0, \text{ and } \sum_{j\neq i_0} q(j|i, a)V_3(j) \leq -\hat{\delta}V_3(i) - 1 \ \forall a \in \mathbf{A}(i), \text{ for all } i \neq i_0.$$

Then, the following assertions hold.

(a) $\psi^f(i) \geq e^{-L(i_0)\sup_{i\in\mathbf{S}} V_3(i)}$ for all $i \in \mathbf{S}$ and $f \in F$, which implies Conditions 3.4.

(b) If in addition $c(i, a) \leq \hat{\delta}$ for all $(i, a) \in \mathbb{K}$, then Conditions 3.5 and 3.4 are satisfied.

Proof. Let $M := \sup_{i\in\mathbf{S}} V_3(i)$. Then,for any $f \in F$, it follows from the proof of Lemma 6.1.5 in [3] that

$$E_i^f(\tau_{i_0}) \leq V_3(i) \leq M \text{ and } E_i^f \left[e^{\hat{\delta}\tau_{i_0}}\right] \leq \hat{\delta}V_3(i) + 1 \leq 1 + \hat{\delta}M < \infty, \text{ for all } i \in \mathbf{S}.$$

Thus, by $\psi^f(i) = E_i^f\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,f(\xi_t))-\rho^f)dt}\right] \geq e^{-\rho^f E_i^f(\tau_{i_0})} \geq e^{-L(i_0)M}$, we see that (a) is true. Obviously, (b) follows from that $\psi^f(i) \leq E_i^f\left[e^{\int_0^{\tau_{i_0}} c(\xi_t,f(\xi_t))dt}\right] \leq E_i^f\left[e^{\hat{\delta}\tau_{i_0}}\right] \leq 1+\hat{\delta}M$. □

We next prove the uniqueness of a solution to (3.28) or (3.27).

Proposition 3.3. Under Conditions 3.1, 3.2, 3.3, 3.4 and 3.5, the followings hold.

(a) The solution (ρ^*, ψ^*) to the RS-AOE (3.27) is unique in $[0, L(i_0)] \times B_1^+(\mathbf{S})$.

(b) For each $f \in F$, the solution (ρ^f, ψ^f) to the multiplicative Poisson equation (3.28) is unique in $[0, L(i_0)] \times B_1^+(\mathbf{S})$.

Proof. (a) In Theorem 3.5, we have shown that (ρ^*, ψ^*) is a solution in $[0, L(i_0)] \times B_{V_0}^+(\mathbf{S})$ to the RS-AOE (3.27) and $(\rho^*, \psi^*) = (\rho^{f^*}, \psi^{f^*})$ for some $f^* \in F$. Under Condition 3.5, it is obvious that $\psi^* \in B_1^+(\mathbf{S})$. Hence, it remains to show that such a solution is unique to the RS-AOE (3.27) in $[0, L(i_0)] \times B_1^+(\mathbf{S})$. To do so, suppose that (ρ, ψ) is an arbitrary solution in $[0, L(i_0)] \times B_1^+(\mathbf{S})$ to the RS-AOE. Since $(\rho, \psi) \in [0, L(i_0)] \times B_1^+(\mathbf{S})$, using a similar argument as in proving Theorem 3.3(b,c) and Theorem 3.5(b) yields that $\rho = \inf_{\pi\in\Pi} J(i, \pi)$, and $\psi(i) = \inf_{\pi\in\Pi_m^d} E_i^\pi\left[e^{\int_0^{\tau_{i_0}}(c(\xi_t,\pi_t)-\rho)dt}\right]$ for all $i \in \mathbf{S}$. It then follows that $\rho = \rho^*$ and $\psi(i) = \psi^*(i)$ for all $i \in \mathbf{S}$.

(b) The proof is similar to those of part (a) above. □

Basing on the uniqueness of a solution to (3.28), we next provide a policy iteration algorithm for computing optimal stationary policies.

The policy iteration algorithm:

1. Pick an arbitrary $f \in F$. Let $n = 0$, and take $f_n := f$.

2. Policy evaluation (by Proposition 3.3(b)): Compute ρ^{f_n} and ψ^{f_n} by solving the following multiplicative Poisson equation

$$\rho\psi(i) = c(i, f_n(i))\psi(i) + \sum_{j\in\mathbf{S}} \psi(j)q(j|i, f_n(i)) \quad \forall\ i \in \mathbf{S},\ \text{with } \psi(i_0) = 1.$$

3. Policy improvement: Obtain a policy $f_{n+1} \in F$ such that, for each $i \in \mathbf{S}$,

$$f_{n+1}(i) := \begin{cases} f_n(i) & \text{when } B_{f_n}(i) = \emptyset \\ a & \text{with any } a \in B_{f_n}(i) \neq \emptyset, \end{cases} \tag{3.29}$$

where

$$B_{f_n}(i) := \{a \in \mathbf{A}(i) | \quad c(i,a)\psi^{f_n}(i) + \textstyle\sum_{j \in \mathbf{S}} q(j|i,a)\psi^{f_n}(j) < \rho^{f_n}\psi^{f_n}(i)\}.$$

if the set $B_{f_n}(i)$ contains more than one action, then we choose any one of them to be $f_{n+1}(i)$.

4. If $f_{n+1} = f_n$ (i.e., $B_{f_n}(i) \equiv \emptyset$),then stop because f_{n+1} is optimal (by Theorems 3.3 and 3.5(c) above). Otherwise, increase n by 1 and return to step 2.

To establish the convergence of this algorithm, $\mathcal{M} := \{S, A(i), c(i,a), q(j|i,a)\}$ needs to be irreducible, which means that $\{\xi_t, t \geq 0\}$ is irreducible under each $f \in F$. Then, we have the following.

Lemma 3.4. Under the conditions in Theorem 3.3, suppose that $\mathbf{S}$ and $\mathbf{A}(i)(i \in \mathbf{S})$ are finite and $\mathcal{M}$ is irreducible. Let $\{f_n\}$ be a sequence obtained by the policy iteration algorithm, then the following assertions hold.

(a) $\rho^{f_{n+1}} \leq \rho^{f_n}$ for all $n \geq 1$.

(b) If $f_{n+1} \neq f_n$ for some $n \geq 0$, then $\rho^{f_{n+1}} < \rho^{f_n}$, and

$$\begin{aligned} \rho^{f_{n+1}} - \rho^{f_n} &= c(i_0, f_{n+1}(i_0))\psi^{f_{n+1}}(i_0) + \sum_{j \in \mathbf{S}} \psi^{f_{n+1}}(j) q(j|i_0, f_{n+1}(i_0)) \\ &\quad - c(i_0, f_n(i_0))\psi^{f_n}(i_0) - \sum_{j \in \mathbf{S}} \psi^{f_n}(j) q(j|i_0, f_n(i_0)). \end{aligned}$$

(c) An optimal policy can be obtained by the algorithm in a finite number of steps.

Proof. By Theorem 5.1 in [43] and Proposition 3.3, we see that (a)-(c) are true. □

In order to get optimal policies for the case of infinite states by finite-approximation, we construct models $\mathcal{M}_n := \{\mathbf{S}_n, \mathbf{A}_n(i), c_n(i,a), q_n(j|i,a)\}(n \geq 1)$ with finite states, where

$$\mathbf{S}_n := \{0, \ldots, n\};\ \mathbf{A}_n(i) := \mathbf{A}(i);\ c_n(i,a) := c(i,a); \text{and}$$
$$q_n(j|i,a) := \begin{cases} q(j|i,a) + \frac{1}{n}\sum_{k>n} q(k|i,a) & \text{for } 0 \leq j \leq n, j \neq i, 0 \leq i \leq n-1 \\ q(i|i,a) & \text{for } j = i \\ q(j|i,a) + \frac{1}{n}\sum_{k>n} q(k|i,a) & \text{for } i = n, 0 \leq j < n. \end{cases} \tag{3.30}$$

for each $i \in \mathbf{S}_n$ and $a \in \mathbf{A}_n(i)$.

Obviously, the transition rates $q_n(j|i,a)(n \geq 1)$ are also conservative and stable. Moreover, if the function V_0 in Condition 3.1 is nondecreasing on $\mathbf{S}$, then Condition 3.1 holds for each model $\mathcal{M}_n$, which is verifies as follows: For each $n \geq 1, i \in \mathbf{S}_n, a \in \mathbf{A}_n(i)$,

$$\begin{aligned} \sum_{j\in\mathbf{S}_n} q_n(j|i,a)V_0(j) &= \sum_{j\in\mathbf{S}_n, j\neq i} [q(j|i,a) + \frac{1}{n}\sum_{k>n} q(k|i,a)]V_0(j) + q(i|i,a)V_0(i) \\ &= \sum_{j\in\mathbf{S}_n} q(j|i,a)V_0(j) + \frac{1}{n}\sum_{k>n} q(k|i,a)[\sum_{j\in\mathbf{S}_n, j\neq i} V_0(j)] \\ &\leq \sum_{j\leq n} q(j|i,a)V_0(j) + \sum_{k>n} q(k|i,a)V_0(k) \\ &\leq -\delta(i)V_0(i) + b_0 I_{\{i_0\}} \qquad \text{(by Condition 3.1).} \end{aligned} \tag{3.31}$$

Thus, in summary, we have the fact below.

Theorem 3.6. Under Conditions 3.1, 3.2 and 3.3, if the functions V_0 and V_1 are nondecreasing on $\mathbf{S}$, then the following assertions hold for each $\mathcal{M}_n(n \geq 1)$.

(a) The solution (ρ_n^*, ψ_n^*) to the RS-AOE (3.32) for $\mathcal{M}_n$ is unique in $[0, L(i_0)] \times$

$B_1^+(\mathbf{S})$:

$$\rho_n^*\psi_n^*(i) = \inf_{a\in\mathbf{A}(i)}\{c_n(i,a)\psi_n^*(i) + \sum_{j\in\mathbf{S}_n}\psi_n^*(j)q_n(j|i,a)\} \quad \forall\ i\in\mathbf{S}_n, \tag{3.32}$$
$$\psi_n^*(i_0) = 1, \rho_n^* \le L_0(i_0),\ \psi_n^*(i) \le V_0(i), \text{for all } i\in\mathbf{S}_n \text{ and } n\ge 1.$$

(b) There exists a $f_n^*\in F$ achieving the minimum in (3.32). Moreover, a policy f in F is optimal for $\mathcal{M}_n$ if and only if $f(i)$ achieves the minimum in (3.32) for all $i\in\mathbf{S}_n$.

(c) If, in addition, $\mathcal{M}$ is irreducible and $\mathbf{A}(i)$ is finite for each $i\in\mathbf{S}$, then an optimal stationary policy f_n^* for $\mathcal{M}_n$ can be obtained by the policy iteration algorithm in a finite number of steps.

Proof. (a)and (b). Since $\mathbf{S}_n$ are finite, it follows from (3.31) that the Conditions 3.1, 3.2, 3.3, 3.4 and 3.5 are satisfied for each $\mathcal{M}_n$. Thus, (a) follows from Theorem 3.5 and Proposition 3.3(a) as well as (3.31), (b) from Proposition 3.3.

(c) First, we show that $\mathcal{M}_n$ is also irreducible for each $n\ge 1$. Indeed, given any stationary policy $f_n\in F$ for model $\mathcal{M}_n$ and $i,j\in\mathbf{S}_n, i\ne j$, we extend f_n to $f\in F$ by letting $f(i) := f_n(i)$ for all $i\in\mathbf{S}_n$ and $f(i) := a_i$ for any $i\notin\mathbf{S}_n$, where $a_i\in\mathbf{A}(i)$ is any fixed action. Then, since $\mathcal{M}$ is irreducible, there exists $K\ge 0$ states $i_k\in\mathbf{S}\setminus\{j\}(k=0,\dots K$, with $i_0=i, i_{K+1}=j)$ such that $q(i_{k+1}|i_k,f(i_k))>0$ for all $k=0,\dots,K$. For the $K+2$ states i_k, if $i_k\in\mathbf{S}_n$ for all $k=0,\cdots,K+1$, then $q(i_{k+1}|i_k,f_n(i_k))>0$ for all $k=0,\cdots,K+1$, which implies that i can reach j under f_n. Otherwise, let $k^* := \min\{k : i_k\notin\mathbf{S}_n\}$. Since $i_0, i_{K+1}\in\mathbf{S}_n$, we have $1\le k^*\le K$ and $i_{k^*+1}\notin\mathbf{S}_n$. Then we have $\{i_0,\dots,i_{k^*-1}\}\subset\mathbf{S}_n$, but $i_{k^*}\notin\mathbf{S}_n$. Thus, by the definition of the transition rates $q_n(\cdot|\cdot,\cdot)$ in (3.30) and $i_{k^*-1}\ne j\in\mathbf{S}_n$, we have

$$q_n(j|i_{k^*-1},f_n(i_{k^*-1})) = q_n(j|i_{k^*-1},f(i_{k^*-1})) \ge \frac{1}{n}q(i_{k^*}|i_{k^*-1},f(i_{k^*-1})) > 0,$$

which, together with $q(i_{k+1}|i_k,f(i_k))>0$ for $k=0,\dots,k^*-1$, implies that i

can reach to j under f_n for the model $\mathcal{M}_n$. Thus, $\mathcal{M}_n$ is irreducible, and so (c) follows from Lemma 3.4. □

Since the sets $\mathbf{A}(i)$ are compact, and so is F. Thus, any sequence $\{f_n^*\}$ in Theorem 3.6(c) has a limit policy (say $\hat{f}^*$) in F, that is, there is a subsequence $\{f_{n_k}^*\}$ of $\{f_n^*\}$ such that $\lim_{k\to\infty} f_{n_k}^*(i) = \hat{f}^*(i)$ for each $i \in \mathbf{S}$.

Theorem 3.7. (Finite-approximation.) Suppose that Conditions 3.1, 3.2 and 3.3 are satisfied, and the V_0 is nondecreasing on $\mathbf{S}$. Then, the followings hold.

(a) A sequence $\{f_n^*\}$ of optimal stationary policies f_n^* exists for $\mathcal{M}_n$, and it has a limit policy $\hat{f}^*$ such that

$$\begin{aligned} \hat{\rho}\hat{\psi}(i) &= c(i,\hat{f}^*(i))\hat{\psi}(i) + \sum_{j\in\mathbf{S}} q(j|i,\hat{f}^*(i))\hat{\psi}(j) \\ &= \inf_{a\in\mathbf{A}(i)} \{c(i,a)\hat{\psi}(i) + \sum_{j\in\mathbf{S}} q(j|i,a)\hat{\psi}(j)\} \quad \forall i \in \mathbf{S}. \end{aligned} \tag{3.33}$$

for some function $\hat{\psi}$ on S such that $\hat{\psi} \le V_0$.

(b) If, in addition, the condition in Proposition 3.2 holds with a nondecreasing V_3 on S and a constant $\hat{\delta} \ge c(i,a)$ on K, then the policy $\hat{f}^*$ in (a) is optimal for the model $\mathcal{M}$.

Proof. (a) Suppose that $\hat{f}^*(i) = \lim_{k\to\infty} f_{n_k}^*(i)$ for all $i \in \mathbf{S}$. Then, by Theorem 3.5, we have $0 \le \rho_{n_k}^* \le L_0(i_0)$ and $\psi_{n_k}^*(i) \le V_0(i)$ for all $k \ge 1$. Thus, the diagonalization arguments ensure the existence of a subsequence $\{n_{k_l}, l \ge 1\}$ of $\{n_k, n \ge 1\}$ and $(\hat{\rho}, \hat{\psi}) \in [0, L_0(i_0)] \times B_{V_0}(\mathbf{S})$ such that $\hat{\psi}(i_0) = 1$ and:

$$\lim_{l\to\infty} f_{n_{k_l}}^*(i) = \hat{f}^*(i),\ \lim_{l\to\infty} \rho_{n_{k_l}}^* = \hat{\rho},\ \lim_{l\to\infty} \psi_{n_{k_l}}^*(i) = \hat{\psi}(i) \le V_0(i)\ \forall i \in \mathbf{S} \tag{3.34}$$

For given $i \in \mathbf{S}$, there exists $\tilde{n}$ such that $i \in \mathbf{S}_n$ for $n \ge \tilde{n}$. Then, Theorem

3.6(a,b) gives $\forall\ l \geq \tilde{n}$

$$\begin{aligned}
\rho^*_{n_{k_l}}\psi^* n_{k_l} &= c_{n_{k_l}}(i, f^*_{n_{k_l}}(i))\psi^*_{n_{k_l}}(i) + \sum_{j\in\mathbf{S}_{n_{k_l}}} \psi^*_{n_{k_l}}(j) q_{n_{k_l}}(j|i, f^*_{n_{k_l}}(i)) \qquad (3.35)\\
&= \inf_{a\in\mathbf{A}(i)}\{c_{n_{k_l}}(i,a)\psi^*_{n_{k_l}}(i) + \sum_{j\in\mathbf{S}_n} \psi^*_{n_{k_l}}(j) q_{n_{k_l}}(j|i,a)\} \qquad (3.36)\\
&\leq c_{n_{k_l}}(i,a)\psi^*_{n_{k_l}}(i) + \sum_{j\in\mathbf{S}_n} \psi^*_{n_{k_l}}(j) q_{n_{k_l}}(j|i,a) \quad \forall\ a\in\mathbf{A}(i).
\end{aligned}$$

On the other hand, since $\psi^*_n \leq V_0$ for all $n \geq 1$, by (3.30), we have

$$\begin{aligned}
&\sum_{j\in\mathbf{S}_n} \psi^*_n(j) q_n(j|i, f^*_n(i)) \qquad (3.37)\\
= &\sum_{j\in\mathbf{S}_n} \psi^*_n(j) q(j|i, f^*_n(i)) + \frac{1}{n}\sum_{k>n} q(k|i, f^*_n(i)) \sum_{0\leq j\leq n, j\neq i} \psi^*_n(j)
\end{aligned}$$

which, together with the monotonicity of V_0 and the following

$$0 \leq \frac{1}{n}\sum_{k>n} q(k|i, f^*_n(i)) \sum_{0\leq j\leq n, j\neq i} \psi^*_n(j) \leq \sum_{k>n} q(k|i, f^*_n(i)) V_0(k) \to 0 \text{ as } n\to\infty,$$

implies that

$$\lim_{n\to\infty}\left[\sum_{j\in\mathbf{S}_n} \psi^*_n(j) q_n(j|i, f^*_n(i))\right] = \sum_{j\in\mathbf{S}} \hat{\psi}^*(j) q(j|i, \hat{f}^*(i)). \qquad (3.38)$$

Thus, by (3.34)-(3.38), we get (3.33).

(b) As the proof of (3.31), we have

$$\hat{\delta} < q_n(i,a), \text{and } \sum_{j\neq i_0} q_n(j|i,a) V_3(j) \leq -\hat{\delta} V_3(i) - 1\ \forall a\in\mathbf{A}(i), \text{ for every } i\neq i_0, n\geq 1,$$

which, together with the proofs of Proposition 3.2 and Theorem 3.5(c), gives $e^{-L(i_0)M} \leq \psi^*_n(i) \leq 1 + \hat{\delta}M$, where $M = \sup_{i\in\mathbf{S}} V_3(i) < \infty$, and so $e^{-L(i_0)M} \leq \hat{\psi}^*(j) \leq 1 + \hat{\delta}M$ for all $i \in \mathbf{S}$. Hence, by (3.33) and Theorem 3.5, we see that (b) is also true. □

3.5 Examples

In this section, we will give two examples, which are used to verify the conditions in this paper and show the difference between the conditions here and those in the previous literature for the risk-sensitive average CTMDPs.

Example 3.2. (*Controlled population processes.*) In a population process, our aim is to minimize the cost of the system caused by birth & death rate and each individual. We regard the population size at any time as the state variable, and suppose that the birth and death parameters can be controlled by a decision maker and denoted by $\lambda_i(a_1)$ and $\mu_i(a_2)$ respectively, which may depend on the system's state i and decision variables (a_1, a_2) taken by the decision maker. When the state of the process is at $i \in \mathbf{S} := \{0, 1, \dots, \}$, the decision maker takes an action $a := (a_1, a_2)$ from a given set $\mathbf{A}(i)$, which may increase or decrease the parameters $\lambda_i(a_1)$ and $\mu_i(a_2)$. On the other hand, because of some possible catastrophe, it is suitable to suppose that a transition from i to 0 may happen at rate $\beta(i)$ for all $i \geq 1$. Choosing any action $a = (a_1, a_2)$ at state i results in some cost denoted by $c(i, a)$. Moreover, the decision maker wishes to minimize the associated risk-sensitive average cost.

We now formulate the controlled population processes as CTMDPs. Obviously, the state space $\mathbf{S} = \{0, 1, \dots\}$ is denumerable; the corresponding transition rates $q(j|i, a)$ are as follows. When there is no population in the system (i.e., $i = 0$), any control of death is unnecessary, and so we set $\mathbf{A}_2(0) := \{0\}$. Thus, we have

$$q(1|0, a) = -q(0|0, a) := \lambda_0(a_1), \qquad \text{for } a = (a_1, a_2) \in \mathbf{A}_1(0) \times \mathbf{A}_2(0),$$

where a_1 denote immigration rates varying in $\mathbf{A}_1(0) := [0, \alpha_1]$ for some constant

$\alpha_1 > 0$. Moreover, for each $i \geq 1, a = (a_1, a_2) \in \mathbf{A}_1(i) \times \mathbf{A}_2(i)$, we have

$$q(j|i,a) = \begin{cases} \lambda_i(a_1) & \text{if } j = i+1, \\ -\lambda_i(a_1) - \mu_i(a_2) - \beta(i) & \text{if } j = i, \\ \mu_i(a_2) & \text{if } j = i-1, i \geq 2, \\ \mu_1(a_2) + \beta(1) & \text{if } j = 0, i = 1, \\ \beta(i) & \text{if } j = 0, i \geq 2 \\ 0 & \text{otherwise.} \end{cases} \tag{3.39}$$

We aim to find conditions imposed on $q(j|i,a)$ in (3.39) and $c(i,a)$, which can ensure the existence of an optimal policy, and thus consider the following sets of hypotheses H_1 and H_2 with given positive constants μ and λ. A characterization of these hypotheses is that their conditions are imposed on the elements of the model and thus can be verified.

H_1 (On controlled birth and death processes with catastrophes [3, p.292]):

(a) $\mathbf{A}(i) := [-\lambda, \lambda] \times [-\mu, \mu]$ for all $i \geq 1$ and $\mu \geq \max\{\lambda, \frac{1}{2}\}$;

(b) $\lambda_i(a_1) := \lambda i + a_1$ for all $i \geq 0$, and $\mu_i(a_2) =: \mu(i+2)^2 + a_2$, for all $i \geq 1$;

(c) $\beta_* := \inf_{i \geq 1} \beta(i) > 0$;

(d) $0 \leq c(i,a) \leq \ln \sqrt{i+2}$ for all $i \geq 0$ and $a \in \mathbf{A}(i)$, and $c(i,a)$ is continuous in $a \in \mathbf{A}(i)$ for each fixed $i \in \mathbf{S}$.

Remark 3.7. The transition and cost rates in the condition H_1 can be unbounded. However, H_1 needs the catastrophe hypothesis (i.e., $\inf_{i \geq 1} \beta(i) > 0$), which is for the usage of Proposition 3.1(b). To remove the catastrophe hypothesis, we need some price of the stochastic monotonicity, and thus modify H_1 as the following H_2, which is for the verification of the conditions in Proposition 3.1(c).

H_2 (On controlled birth and death processes without any catastrophe [3, p.292]):

(a) $\mathbf{A}_1(0) = [0, \alpha_1], \mathbf{A}(i) := [-\lambda, \lambda] \times [\mu, \mu]$ for all $i \geq 1$, where $\mu \geq \max\{\lambda, \frac{1}{2}\}$;

(*b*) $\lambda_i(a_1) := \lambda i + a_1$ for $i \geq 0$, and $\mu_i(a_2) =: \mu(i+2)^2 + a_2$, for all $i \geq 1$;

(*c*) $\beta(i) = 0$ for all $i \geq 1$;

(*d*) $0 \leq c(i,a) \leq \ln\sqrt{i+2}$ for all $(i,a) \in K$, and $c(i,a)$ is continuous in $a \in \mathbf{A}(i)$ for each fixed $i \in \mathbf{S}$;

(*e*) $\inf_{a' \in \mathbf{A}(i+1)} c(i+1,a') \geq c(i,a)$ for all $(i,a) \in K$, which implies that $c(i,f(i))$ is increasing in $i \in \mathbf{S}$ for any given $f \in F$.

Proposition 3.4. Under one of H_1 and H_2, conditions 3.1, 3.2, 3.3 and 3.4 are satisfied. Thus, an optimal stationary policy exists for Example 3.2.

Proof. (a) Under H_1, in order to verify the assumptions, let $i_0 := 0$, and

$$V_0(i) := (i+2)^2, \delta(i) := \ln\sqrt{i+2},\ V_1(i) := (i+2)^4,\ V_2(i) := \frac{1}{\beta_*} \text{ for } i \neq 0, V_2(0) := 0.$$

Then, using the condition in H_1, a directive calculation gives, for all $i \geq 1$ and $(a_1,a_2) \in \mathbf{A}(i)$,

$$\begin{aligned}
\sum_{j\in\mathbf{S}} q(j|0,a)V_0(j) &= 4a_1 \leq -\delta(0)V_0(0) + 8\mu + 4b_1, && (3.40)\\
\sum_{j\in\mathbf{S}} q(j|i,a)V_0(j) &\leq -\frac{\mu}{2}(i+2)V_0(i) \leq -\delta(i)V_0(i); && (3.41)\\
\sum_{j\in\mathbf{S}} q(j|i,a)V_1^2(j) &\leq 192(\lambda+\mu)(i+2)^8 = 192(\lambda+\mu)V_1^2(i); && (3.42)\\
\sum_{j\neq 0} q(j|i,a)V_2(j) &= -\frac{\beta(i)}{\beta_*} \leq -1. && (3.43)
\end{aligned}$$

Thus, since $q^*(i) \leq (\lambda+\mu)(i+2)^2 \leq (\lambda+\mu)V_0(i)$, Condition 3.1 follows from (3.40)-(3.41). From the descriptions of the example and H_1, we see that Condition 3.3 is satisfied, and (3.42) implies Condition 3.2. Thus, the rests need to verify Condition 3.4. Indeed, since $\inf_{a\in\mathbf{A}(i)} q(i,a) = \inf_{a\in\mathbf{A}(i)}(\lambda i + a_1 + \mu(i+2)^2 + a_2 + \beta(i)) \geq 3\mu > 0$ (by H_1) for each $i \neq 0$, by (3.43) and Proposition 3.1(b), we know Condition 3.4 is satisfied.

Under H_2, as the arguments for H_1, we see that Conditions 3.1-3.2 are all satisfied. Moreover, since the birth and death processes without any catastrophe (by $H_2(c)$) are stochastic monotone and the $H_2(e)$ implies that $c(i, f(i))$ is nondecreasing in $i \in \mathbf{S}$ (for any given stationary policy f), Condition 3.4 follows from Proposition 3.1(c). Thus , Theorem 3.5 gives the desirable result. □

Remark 3.8. Under H_1 or H_2, take any policy stationary policy f with $f(0) = (0, 0)$ (or $f(1) = (-\mu, -\lambda)$. Then, we have $q(0|0, f(0)) = 0$ (or $q(1|1, f(1)) = 0$), and so the state "0" (or "1") is absorbing under the policy f, while the state process under every stationary policies has been assumed to be ergodic in [43, 74, 73, 102]. Of course, if the sets $\mathbf{A}(0)$ and $\mathbf{A}(i)(i \geq 1)$ are taken to be compact subsets of $(0, b_1]$ and $(-\lambda, \lambda) \times (-\mu, \mu)$ respectively, then the $\mathcal{M}$ (i.e., process $\{x_t, t \geq 0\}$) is irreducible under any stationary policy.

To illustrate the calculation of optimal stationary policies for the case of infinite states by the policy iteration algorithm, we consider the following conditions H_3.

H_3 (On (On controlled irreducible birth and death processes with catastrophes [3, p.292])):

- (a) $\mathbf{A}_1(0) := [\alpha_2, \alpha_3]$ for constants $\alpha_3 > \alpha_2 > 0$, and $\mathbf{A}(i) := [0, \lambda] \times [0, \mu]$ for $i \geq 1$;
- (b) $\lambda_i(a_1) := \lambda i + a_1$ for all $i \geq 0$, $\mu_i(a_2) =: \mu i + a_2$, for all $i \geq 1$, and $\mu \geq \lambda$;
- (c) $\beta(i) \geq \max\{2 + 3(\mu + \lambda) + L, 4\sqrt{\ln(1 + i)} + 2\lambda - \mu\}$ for $i \geq 1$, with a constant $L > 0$;
- (d) $0 \leq c(i, a) \leq \min\{L, \frac{1}{4}(2 + 4\mu + \lambda + L)\}$ for all $i \geq 0$ and $a \in \mathbf{A}(i)$, and $c(i, a)$ is continuous in $a \in \mathbf{A}(i)$ for each fixed $i \in \mathbf{S}$.

Proposition 3.5. Under H_3 for Example 3.2, an optimal policy exists and can be obtained as a limit policy of $\{f_n^*\}$ of optimal policies f_n^* for the models $\mathcal{M}_n$.

Proof. Let $i_0 := 0, V_0(i) := 1+i, \delta(i) := \frac{1}{4}(\beta(i)+\mu-2\lambda), V_1(i) := (1+i)^2, V_3(i) := \frac{i}{1+i}$ for all $i \geq 0$, and $\hat{\delta} := L$. Then, V_0 and V_3 are nondecreasing on $\mathbf{S}$, and $q_* = \inf_{a\in\mathbf{A}(i),i\geq 1} q(i,a) = \lambda + \mu > \hat{\delta}$. Moreover, simple calculations give that $c(i,a) \leq \delta(i) \leq \sqrt{\ln V_0(i)}(a \in \mathbf{A}(i))$ and

$$\begin{aligned}
\sum_{j\in\mathbf{S}} q(j|i,a)V_0(j) &\leq -\delta(i)V_0(i) + 3\alpha_3 I_{\{0\}}(i), \quad \text{for } i \in \mathbf{S}, \\
\sum_{j\in\mathbf{S}} q(j|i,a)V_1^2(j) &= : p_1 + p_2 i + p_3 i^2 + p_4 i^3 + p_4 i^4 \leq K_1 V_1^2(i) + K_2, \quad \text{for } i \in \mathbf{S}, \\
\sum_{j\in\mathbf{S}} q(j|0,a)V_3(j) &= \frac{a_1}{2} \leq \alpha_3, \ \sum_{j\neq 0} q(j|i,a)V_3(j) \leq -LV_3(i) - 1 \text{ for } i \geq 1,
\end{aligned}$$

where the constants $p_1, p_2, p_3, p_4, K_1, K_2$ are determined by the given λ and μ. Thus, from the inequalities above and Proposition 3.2 we see that all conditions in Theorem 3.7(b) are satisfied, and thus the result follows from Theorem 3.7. □

Example 3.3. (*The stochastic logistic process with immigration [3, p.307].*) This is a birth and death process with a finite state space $\mathbf{S} := \{0, 1, \ldots, N\}$, the birth rate $\lambda_i(a_1) := \lambda i(1 - \frac{i}{N})$ for all $i \geq 1$, the death rates $\mu_i(a_2) := \mu i(1 + \frac{a_2 i}{N})$ (for all $i \geq 0$) with parameters a_2, and immigration rates a_1 when there is no population, where λ and μ are given positive constants. Suppose that the parameters (a_1, a_2) may be changed in the set $\{0, 1, \ldots, \bar{a}\} \times \{0, 1, \ldots, \mu\}]$ for some $\bar{a} \geq 1, \mu \geq 1$, and any change of $a = (a_1, a_2)$ at state i results in some cost $c(i, a)$. We wish to minimize the associated risk-sensitive average cost.

Obviously, the model of CTMDPs for the stochastic logistic process with immigration is as follows: $\mathbf{S} = \{0, 1, \ldots, N\}$, $\mathbf{A}(0) = \{0, 1, \ldots, \bar{a}\} \times \{\mu\}, \mathbf{A}(i) =$

$\{\lambda\} \times \{0, 1, \ldots, \mu\}$ for all $i \geq 1$, the transition rates $q(j|i, a)$

$$q(j|i,a) = \begin{cases} \lambda i(1 - \frac{i}{N}) & \text{if } j = i+1, i \geq 1 \\ -\lambda i(1 - \frac{i}{N}) - \mu i(1 + \frac{a_2 i}{N}) & \text{if } j = i \geq 1 \\ \mu i(1 + \frac{a_2 i}{N}) & \text{if } j = i-1, i \geq 1, \\ -a_1 & \text{if } j = 0, i = 0, \\ a_1 & \text{if } j = 1, i = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.44}$$

Proposition 3.6. If $0 \leq c(i,a) < \frac{1}{3}(\mu - \lambda + \frac{\lambda}{N}) \leq \ln 2$, $\mu > \lambda$, and $c(i,a)$ is continuous in $a \in \mathbf{A}(i)$ for each $i \in \mathbf{S}$, then Example 3.3 has an optimal stationary policy, which can be obtained by the policy iteration algorithm in a finite number of iteration steps.

Proof. Let

$$V_0(i) := i + 2, V_1(i) := (N+2)(i+2), \delta(i) \equiv \frac{1}{3}(\mu - \lambda + \frac{\lambda}{\mu}) \quad \forall\, 0 \leq i \leq N.$$

Then, we have $V_0^2(i) \leq V_1(i)$ for all $0 \leq i \leq N$, and

$$\sum_{j \in \mathbf{S}} q(j|0,a) V_0(j) = a_1 \leq -\delta(0) V_0(0) + \frac{2}{3}(\mu - \lambda + \frac{\lambda}{\mu}) + \bar{a}; \tag{3.45}$$

$$\sum_{j \in \mathbf{S}} q(j|i,a) V_0(j) = -\mu i(1 + \frac{a_2 i}{N}) + \lambda i(1 - \frac{i}{N}) \leq -\delta(i) V_0(i); \tag{3.46}$$

$$\begin{aligned} \sum_{j \in \mathbf{S}} q(j|i,a) V_1^2(j) &= (N+2)^2[-\mu i(1 + \frac{a_2}{N} i)(2i+3) + \lambda i(1 - \frac{i}{N})(4i+8)] \\ &\leq (N+2)^2 \lambda i(4i+8) \leq 4\lambda V_1^2(i). \end{aligned} \tag{3.47}$$

Thus, since $q^*(i) \leq N(1 + \lambda + \mu)^2$ for all $0 \leq i \leq N$, Condition 3.1 follows from (3.45)-(3.46). From the descriptions of this example and (3.44), we see that Condition 3.3 is satisfied, and (3.47) implies Condition 3.2. Moreover, Condition 3.4 follows from Proposition 3.1(a) and the finiteness of the states space. □

Remark 3.9. In the conditions H_1 and H_2 for Example 3.2, the cost and transition rates are allowed to be unbounded. Moreover, if the $c(i,a)$ in $H_1(\mathrm{d})$ is unbounded (for example, $c(i,a) := \sqrt{\ln(i+1+\frac{|a_1|+|a_2|}{\mu+\lambda})}$, then the smallness condition (stronger than the standard boundedness condition) on the costs in [43, 74] is not satisfied. If taking $c(i,a) = \sqrt{\ln[1+\frac{|a_1|+|a_2|}{(\mu+\lambda)(i+1)}]}$ which satisfied $H_1(\mathrm{d})$, then $\liminf_{i\to\infty} \inf_{a\in\mathbf{A}(i)} c(i,a) = 0$. But the near-monotone condition in [73] (i.e., $\liminf_{i\to\infty} \inf_{a\in\mathbf{A}(i)} c(i,a) > \inf_{f\in F} J(i,f)$) fails to hold; Furthermore, the uniform boundedness hypothesis on the transition rates in [43, 74, 73, 102] fail to hold for $H_1(\mathrm{b})$.

4 Risk-sensitive gradual-impulse CTMDP

4.1 Introduction

In this chapter, we consider the gradual-impulse control problem of continuous-time Markov decision processes, where the system performance is measured by the expectation of the exponential utility of the total cost. We prove, under very general conditions on the system primitives, the existence of a deterministic stationary optimal policy out of a more general class of policies. Policies that we consider allow multiple simultaneous impulses, randomized selection of impulses with random effects, relaxed gradual controls, and accumulation of jumps. After characterizing the value function using the optimality equation, we reduce the continuous-time gradual-impulse control problem to an equivalent simple discrete-time Markov decision process, whose action space is the union of the sets of gradual and impulsive actions.

There is no lack of situations, where an action can affect the state of the controlled process instantaneously. For example, in a Susceptible-Infected-Recovered (SIR) epidemic model, the controller elaborates the immunization policy, affecting the transition rate from the susceptibles to the infectives, as well as the isolation policy, which reduces instantaneously the number of infectives. Let us formulate another simple example, which contains some features motivating this chapter.

Example 4.1. A rat (or intruder) may invade the kitchen. For each time unit it remains alive in the "kitchen", a constant cost of $l \geq 0$ is incurred. The rat spends an exponentially distributed amount of time with mean $\frac{1}{\mu} > 0$ in the kitchen, and then goes outside and settles down in another house (and thus never returns). When the rat is in the kitchen, the housekeeper (defender) can decide to shoot at it, with a chance of hitting and killing the rat being $p \in (0, 1)$. If the rat dodged, it remains in the kitchen. Each bullet costs $C > 0$. Assume that the successive shootings are independent.

Let us mention some features in the above example. "Shoot" is an impulse. The location of the rat is the state. The effect of an impulse on the post-impulse state is random, as the shooting may be dodged. It is costly for each time unit the rat is present in the kitchen. Suppose the cost of impulse is relatively low. It can happen that after one impulse, if the rat is still alive and in the kitchen, then it is reasonable to immediately shoot again. This means, one should allow multiple impulses at a single time moment in this problem. We will return to this problem in Example 4.2 below, which demonstrates the situations when applying only one impulse is insufficient for optimality.

4.2 Model description and problem statement

4.2.1 System primitives of the gradual-impulse control problem

We describe the primitives of the model as follows. Because there are two kind of spaces here containing both continuous and discrete time cases, we denote all of the gradual control model notations with index G (the same as our previous notations for CTMDP, see Chapter 2) and the impulsive control model with index I. Therefore, the space of gradual controls is $\mathbf{A}^G$, and the space of impulsive controls is $\mathbf{A}^I$.

If the current state is $x \in \mathbf{S}$, and an impulsive control $b \in \mathbf{A}^I$ is applied, then the state immediately following this impulse obeys the distribution given by $Q(dy|x,b)$, which is a stochastic kernel from $\mathbf{S} \times \mathbf{A}^I$ to $\mathcal{B}(\mathbf{S})$. Finally, given the current state $x \in \mathbf{S}$, the cost rate of applying a gradual control $a \in \mathbf{A}^G$ is $c^G(x,a)$ and the cost of applying an impulsive control $b \in \mathbf{A}^I$ is $c^I(x,b,y)$, where c^G and c^I are $[0,\infty)$-valued measurable functions on $\mathbf{S} \times \mathbf{A}^G$ and $\mathbf{S} \times \mathbf{A}^I \times \mathbf{S}$, respectively.

Throughout the chapter, we assume that both action space $\mathbf{A}^G$ and $\mathbf{A}^I$ are compact Borel spaces. It is without loss of generality to assume $\mathbf{A}^G$ and $\mathbf{A}^I$ as two disjoint compact subsets of a Borel space $\tilde{\mathbf{A}}$, for otherwise, one can consider

$\mathbf{A}^G \times \{G\}$ instead of $\mathbf{A}^G$ and $\mathbf{A}^I \times \{I\}$ instead of $\mathbf{A}^I$ and $\tilde{\mathbf{A}} = \mathbf{A}^G \times \{G\} \bigcup \mathbf{A}^I \times \{I\}$. Furthermore, we assume that

$$\sup_{a \in \mathbf{A}^G} c^G(x, a) < \infty, \ \forall x \in \mathbf{S}. \tag{4.1}$$

In what follows, we will not make specific reference to this assumption.

The system dynamics in the concerned gradual-impulse control problem can be described as follows. In absence of impulses, the system is just a controlled Markov pure jump process in the state space $\mathbf{S}$, where the (gradual) control, selected from $\mathbf{A}^G$, acts on the local characteristics of the process, leading to natural jumps. This is conveniently described as a marked point process, which consists of the pairs of subsequent jump moments and the the post-jump states (marks). The mark space is thus $\mathbf{S}$. We would still describe the system in the concerned gradual-impulse control problem using a marked point process. However, when the decision maker is allowed to apply a finite or countably infinite sequence of impulses from $\mathbf{A}^I$ at a single time moment, and each impulse results in a post-impulse state, there would be a sequence of states in $\mathbf{S}$ at a single time moment. Moreover, the order of the impulses and their resulting states is also relevant. Therefore, the marked point process we use now is in an enlarged mark space. More precisely, each mark contains a sequence of impulses applied at the same time moment, the state before the impulses are applied, and all the states resulted by these impulses. Each jump moment is either triggered by an impulse (or a sequence of impulses), or by a natural jump. A mark in this marked point process is referred to as an intervention. This term is naturally understandable when the mark consists of impulses. Having said so, we will also allow that an "intervention" does not contain any impulse or say an empty sequence of impulses. This appears when the decision maker chooses not to apply any impulse immediately after a natural jump. In the rest of this section, following the method of [28], we will elaborate this idea and describe rigorously the concerned continuous-time gradual-impulse control problem. To this end, we

will firstly state the precise definition of an intervention in the next section.

4.2.2 Definition and interpretation of an intervention

At the beginning of an intervention, the decision marker chooses whether to apply an impulse, and which one to apply. If the current state is $x \in \mathbf{S}$, and after an impulse $b \in \mathbf{A}^I$ is chosen, the new state say $y \in \mathbf{S}$ is instantaneously realized, following the distribution $Q(dy|x,b)$. Then based on x, b, y, the decision maker will choose the next impulse, if any at all, and so on. To be consistent, a cemetery point $\Delta \notin \mathbf{A}^I, \mathbf{S}$ is artificially fixed, which is chosen when the decision maker decides not to apply any more impulse at the current time, and it leads to the post-impulse state, also denoted as Δ, which is absorbing, i.e., $Q(\Delta|\Delta,\Delta) \equiv 1$. Therefore, an intervention is a sequential decision process. More precisely, an intervention can be regarded as a trajectory or sample path of the following DTMDP, which we refer to as the "intervention" DTMDP model, to distinguish it from several other DTMDP models to appear subsequently.

Definition 4.1. The intervention DTMDP model is specified by the following tuple $\{\mathbf{S}_\Delta, \mathbf{A}^I_\Delta, Q\}$, which are defined in terms of the primitives of the gradual-impulse control problem given in Subsection 4.2.1.

- The state space is $\mathbf{S}_\Delta := \mathbf{S} \bigcup \{\Delta\}$, where Δ is a cemetery point not belonging to $\mathbf{S}$ or $\mathbf{A}^I$.
- The action space is $\mathbf{A}^I_\Delta := \mathbf{A}^I \bigcup \{\Delta\}$.
- The one-step transition probability from $\mathbf{S}_\Delta \times \mathbf{A}^I_\Delta$ to $\mathcal{B}(\mathbf{S}_\Delta)$ is $Q(dy|x,b)$, where we have accepted that $Q(\{\Delta\}|x,b) := 1$ if $x = \Delta$ or $b = \Delta$.

Let the initial distribution in the intervention DTMDP be always concentrated on $\mathbf{S}$. Then its canonical sample space is

$$\mathbf{Y} := \left(\bigcup_{k=0}^{\infty} \mathbf{Y}_k \right) \bigcup (\mathbf{S} \times \mathbf{A}^I)^\infty,$$

where for each $\infty > k \geq 1$

$$\mathbf{Y}_k := (\mathbf{S} \times \mathbf{A}^I)^k \times (\mathbf{S} \times \{\Delta\}) \times (\{\Delta\} \times \{\Delta\})^\infty,$$

and $\mathbf{Y}_0 := (\mathbf{S} \times \{\Delta\}) \times (\{\Delta\} \times \{\Delta\})^\infty$. Here, if $y \in \mathbf{Y}_k$, $\infty > k \geq 0$, then there are k impulses applied in the intervention y. Similarly, if $y \in (\mathbf{S} \times \mathbf{A}^I)^\infty$, then there are infinitely many impulses applied in the intervention y. Now we give the following definition.

Definition 4.2. An intervention is an element of $\mathbf{Y}$.

In other words, $\mathbf{Y}$ defined above is the space of all interventions. It will be the mark space of the marked point process $\{(T_n, Y_n)\}$ introduced in the next subsection.

With the notations introduced above, we now reiterate, more rigorously compared to the one in the beginning of this subsection, the interpretation of an intervention as follows. Given the current state $x \in \mathbf{S}$, if the controller decides to use Δ, then it means, no more impulse is used at this moment, and the intervention DTMDP is absorbed at Δ; if the controller decides to use an impulse $b \in \mathbf{A}^I$, then the post-impulse state follows the distribution $Q(dy|x, b)$. At the next post-impulse state y, if $y = \Delta$, then the only decision is Δ; if $y \neq \Delta$, then the controller either decides to use no impulse, leading to the next post-impulse state Δ, or to use impulse b', leading to the next post-impulse state, which follows the distribution given by $Q(\cdot|y, b')$, and so on. In other words, an intervention consists of a state and a finite or countable sequence of pairs of impulsive actions and the associated post-impulse states. In particular, no impulse is applied in an intervention if the intervention belongs to $\mathbf{Y}_0$, see Figure 1 and its caption for an example. Let

$$\mathbf{Y}^* := \mathbf{Y} \setminus \mathbf{Y}_0 = \left(\bigcup_{k=1}^{\infty} \mathbf{Y}_k \right) \bigcup (\mathbf{S} \times \mathbf{A}^I)^\infty$$

be the set of interventions, where some impulses are applied.

In an intervention, locally, the selection of impulses (including the "pseudo" impulse Δ) from $\mathbf{A}_{\Delta}^{I}$ is governed by a strategy in the intervention DTMDP model. This adverb "locally" is understood in comparison with the definition of a policy for the gradual-impulse control problem, as given in Definition 4.3 below, which governs the selection of impulsive controls as well as gradual controls, and is thus "global". Let Ξ be the set of (possibly randomized and history-dependent) strategies σ in the intervention DTMDP. The way how a strategy in the intervention DTMDP model is incorporated into a policy in Definition 4.3 below is through its strategic measure. We recall the definition of a strategic measure in a DTMDP model in Definition B.1. Let $\beta^{\sigma}(\cdot|x)$ denote the corresponding strategic measure of a strategy σ of the intervention DTMDP, given the initial state $x \in \mathbf{S}$. By the Ionescu-Tulcea theorem, see e.g., Proposition C.10 in [57], the mapping $x \in \mathbf{S} \to \beta^{\sigma}(\cdot|x)$ is measurable. Let $\mathcal{P}^{\mathbf{Y}}$ be the collection of all such stochastic kernels generated by some strategy $\sigma \in \Xi$, and

$$\mathcal{P}^{\mathbf{Y}}(x) := \{\beta^{\sigma}(\cdot|x) : \sigma \in \Xi\}$$

for each state $x \in \mathbf{S}$. Let

$$\mathcal{P}^{\mathbf{Y}^*} := \{\beta(\cdot|\cdot) \in \mathcal{P}^{\mathbf{Y}} : \ \beta(\mathbf{Y}^*|x) = 1, \ \forall\, x \in \mathbf{S}\},$$

and for each $x \in \mathbf{S}$,

$$\mathcal{P}^{\mathbf{Y}^*}(x) := \{\beta(\cdot|x) : \beta(\cdot|\cdot) \in \mathcal{P}^{\mathbf{Y}}, \ \beta(\mathbf{Y}^*|x) = 1\}.$$

4.2.3 Construction of the controlled processes

Let us now describe the promised marked point process $\{(T_n, Y_n)\}_{n=1}^{\infty}$ for the system dynamics of the concerned gradual-impulse control problem, where the mark space is the space of interventions. Then the continuous-time controlled

process $\{\xi_t\}_{t\geq 0}$ is defined based on the marked point process.

Let

$$\begin{aligned}
&\mathbf{Y}_\Delta := \mathbf{Y} \bigcup \{\Delta\},\\
&\Omega_0 := \mathbf{Y} \times (\{0\} \times \mathbf{Y}) \times (\{\infty\} \times \{\Delta\})^\infty,\\
&\Omega_n := \mathbf{Y} \times (\{0\} \times \mathbf{Y}) \times ((0,\infty) \times \mathbf{Y})^n \times (\{\infty\} \times \{\Delta\})^\infty, \forall\ n = 1, 2, \ldots.
\end{aligned}$$

The canonical space Ω is defined as

$$\Omega := \left(\bigcup_{n=0}^{\infty} \Omega_n \right) \bigcup \left(\mathbf{Y} \times ((0,\infty) \times \mathbf{Y})^\infty \right)$$

and is endowed with its Borel σ-algebra denoted by $\mathcal{F}$. The following generic notation of a point in Ω will be in use: $\omega = (y_0, \theta_1, y_1, \theta_2, y_2, \ldots)$. where we recall y_i is the intervention and θ_i is the sojourn time between two interventions. Below, unless stated otherwise, $x_0 \in \mathbf{X}$ will be a fixed notation as the initial state of the gradual-impulse control problem. Then we put

$$y_0 := (x_0, \Delta, \Delta, \ldots),\ \theta_1 \equiv 0. \tag{4.2}$$

The sequence of $\{\theta_n\}_{n=1}^\infty$ represents the sojourn times between consecutive interventions. Here $\theta_1 = 0$ corresponds to that we allow the possibility of applying impulsive control at the initial time moment, c.f. (4.5) below. For each $n = 0, 1, \ldots,$ let

$$h_n := (y_0, \theta_1, y_1, \theta_2, y_2, \ldots \theta_n, y_n) = (y_0, 0, y_1, \theta_2, y_2, \ldots \theta_n, y_n),$$

where the second equality holds because $\theta_1 \equiv 0$, see (4.2). The collection of all such fragmental histories h_n is denoted by $\mathbf{H}_n$. Let us introduce the coordinate

mappings:

$$Y_n(\omega) = y_n,\ \forall\ n \geq 0;\ \Theta_n(\omega) = \theta_n,\ \forall\ n \geq 1.$$

The sequence $\{T_n\}_{n=1}^{\infty}$ of $[0,\infty]$-valued mappings is defined on Ω by $T_n(\omega) := \sum_{i=1}^{n} \Theta_i(\omega) = \sum_{i=1}^{n} \theta_i$ and $T_\infty(\omega) := \lim_{n\to\infty} T_n(\omega)$. Let $H_n := (Y_0, \Theta_1, Y_1, \ldots, \Theta_n, Y_n)$. Finally, we define the controlled process $\left\{\xi_t\right\}_{t\in[0,\infty)}$:

$$\xi_t(\omega) = \begin{cases} Y_n(\omega), & \text{if } T_n \leq t < T_{n+1} \text{ for } n \geq 1; \\ \Delta, & \text{if } T_\infty \leq t, \end{cases}.$$

It is convenient to introduce the random measure μ of the marked point process $\{(T_n, Y_n)\}_{n=1}^{\infty}$ on $(0,\infty) \times \mathbf{Y}$:

$$\mu(dt \times dy) = \sum_{n\geq 2} I_{\{T_n<\infty\}} \delta_{(T_n,Y_n)}(dt \times dy),$$

where the dependence on ω is not explicitly indicated. Let $\mathcal{F}_t := \sigma\{H_1\} \vee \sigma\{\mu((0,s] \times B) :\ s \leq t, B \in \mathcal{B}(\mathbf{Y})\}$ for $t \in [0,\infty)$.

We use the following notation in next definition. For each $y = (x_0, b_0, x_1 \ldots) \in \mathbf{Y}$

$$\bar{x}(y) := x_k$$

if $\infty > k = 0, 1, \ldots$ is the unique integer such that $y \in \mathbf{Y}_k$ (if $k \geq 1$, then $\bar{x}(y)$ is the state after the last impulse in the intervention y); if such an integer k does not exist, then $y \in (\mathbf{X} \times \mathbf{A}^I)^\infty$ and

$$\bar{x}(y) := \Delta.$$

That previous equality corresponds to that we kill the process after an infinite number of impulses was applied at a single time moment. An example of a

trajectory of the system dynamics in the gradual impulse control problem is displayed in Figure 1.

Definition 4.3. A policy is a sequence $u = \{u_n\}_{n=0}^{\infty}$ such that $u_0 \in \mathcal{P}^{\mathbf{Y}}$ and, for each $n = 1, 2, \ldots,$

$$u_n = \left(\Phi_n, \Pi_n, \Gamma_n^0, \Gamma_n^1\right),$$

where Φ_n is a stochastic kernel on $(0, \infty]$ given $\mathbf{H}_n$, Π_n is a stochastic kernel on $\mathbf{A}^G$ given $\mathbf{H}_n \times (0, \infty)$ such that $\Phi_n(\{\infty\}|h_n) = 1$ if $y_n \in (\mathbf{S} \times \mathbf{A}^I)^{\infty}$, Γ_n^0 is a stochastic kernel on $\mathbf{Y}$ given $\mathbf{H}_n \times (0, \infty) \times \mathbf{S}$ satisfying $\Gamma_n^0(\cdot|h_n, t, x) \in \mathcal{P}^{\mathbf{Y}}(x)$ for each $h_n \in \mathbf{H}_n$ and $x \in \mathbf{S}$ and $t \in (0, \infty)$; and Γ_n^1 is a stochastic kernel on $\mathbf{Y}$ given $\mathbf{H}_n$ satisfying $\Gamma_n^1(\cdot|h_n) \in \mathcal{P}^{\mathbf{Y}^*}(\bar{x}(y_n))$ for each $h_n \in \mathbf{H}_n$. (The above conditions apply when $y_n \neq \Delta$; otherwise, all the values of $\Phi_n(\cdot|h_n)$, $\Pi_n(\cdot|h_n, t)$, $\Gamma_n^0(\cdot|h_n, t, \cdot)$ are immaterial and may be put arbitrarily.)

The set of policies is denoted by $\mathcal{U}$.

Let us provide an interpretation of how a policy u acts on the system dynamics. Roughly speaking, an intervention is over as soon as the (possibly empty) sequence of simultaneous impulses is over. Given that the nth intervention is over, the kernel Φ_n specifies the conditional distribution of the planned time until the next impulse (or next sequence of impulses). The (conditional) distribution of the time until the next natural jump (if there were no interventions before it) is the non-stationary exponential distribution with rate $\int_{\mathbf{A}^G} q_{\bar{x}(Y_n)}(a)\Pi_n(da|H_n, t)$. In other words, Π_n is the relaxed gradual control. Given the nth intervention is over, the next intervention is triggered by either the next planned impulse or the next natural jump; in the former case, the new intervention has the distribution given by Γ_n^1, and in the latter case the new intervention has the distribution given by Γ_n^0. This interpretation will be seen consistent with (4.3) and (4.4) below, where one can see how a policy u acts on the conditional law of the marked point process $\{(T_n, Y_n)\}_{n=1}^{\infty}$. See also the caption of Figure 1.

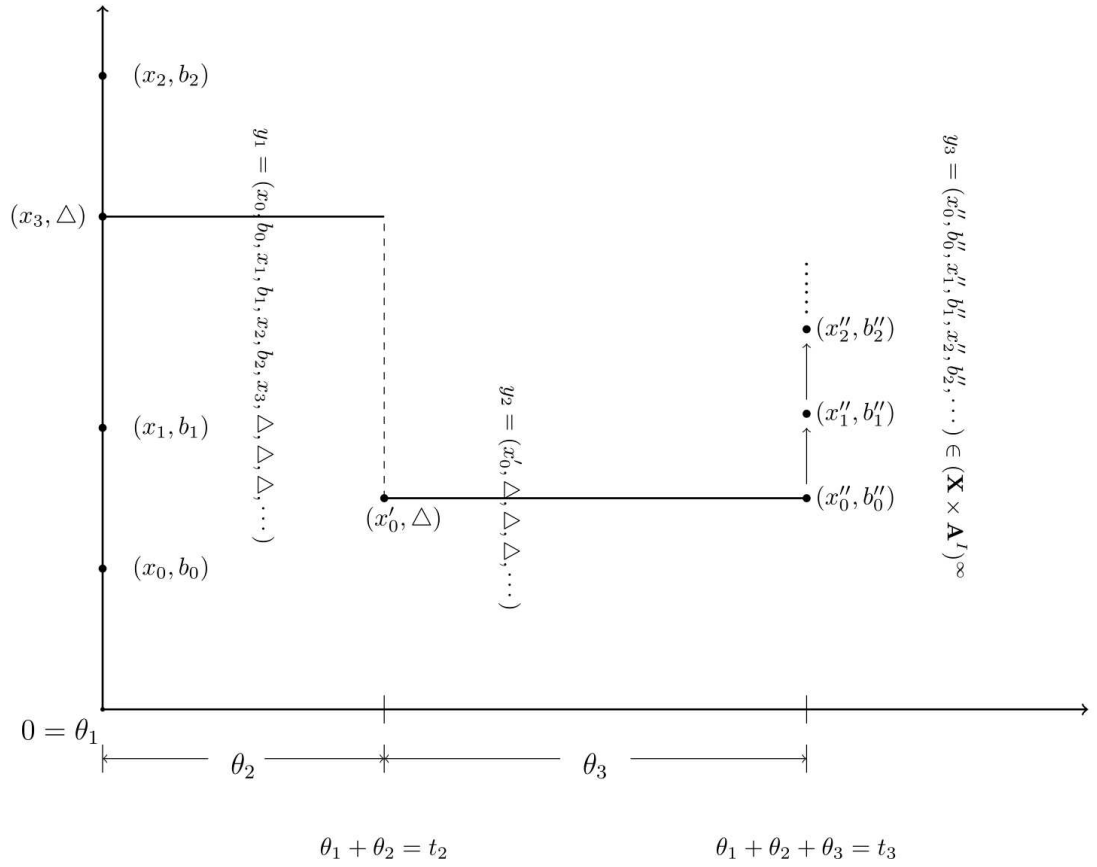

Figure 1: Illustration of the system dynamics in the gradual-impulse control problem, and how the policy acts on the system dynamics. Here $\mathbf{X} = [0, \infty)$. The second coordinate indicates the impulse (including the "pseudo" impulse Δ) used at that state, which is recorded in the first coordinate. At the initial time $t = \theta_1 \equiv 0$, three impulses are applied in turn. The first jump in the indicated sample path of the marked point process $\{(T_n, Y_n)\}_{n=1}^{\infty}$ takes place at $t_2 = \theta_2$. It is triggered by a natural jump because $x_0' \neq x_3$. Along the displayed sample path, the system state remains to be x_3 before the first jump of the marked point process. The second jump of the marked point process is triggered by a planned (or say active) impulse, because $x_0'' = x_0'$. Infinitely many impulses are applied at $t_3 = t_2 + \theta_3$, so that the process is "killed" after the infinitely many impulses at t_3, i.e., $\omega = (y_0, 0, y_1, \theta_2, y_2, \theta_3, y_3, \infty, \Delta, \infty, \Delta, \dots)$. Note also that, under the policy $u = \{u_n\}_{n=0}^{\infty}$ in Definition 4.3, $y_1 \in \mathbf{Y}_3$ is a realization from the distribution $u_0(\cdot|x_0)$, $\bar{x}(y_1) = x_3$; $y_2 \in \mathbf{Y}_0$ is a realization from the distribution $\Gamma_1^0(\cdot|h_1, \theta_2, x_0')$ as the jump at t_2 is triggered by a natural jump, $\bar{x}(y_2) = x_0'$; and $y_3 \in (\mathbf{X} \times \mathbf{A}^I)^\infty$ is a realization from the distribution $\Gamma_2^1(\cdot|h_2)$ as the jump at t_3 is not triggered by a natural jump, $\bar{x}(y_3) = \Delta$.

Suppose a policy $u = \{u_n\}_{n=0}^{\infty}$ is fixed. Let us now present the conditional law of the marked point process $\{(T_n, Y_n)\}_{n=1}^{\infty}$ under the policy u, which determines the underlying probability measure $\mathrm{P}_{x_0}^u$ on $(\Omega, \mathcal{F})$, where $x_0 \in \mathbf{X}$ is the fixed initial state of the system dynamic. For brevity, we introduce the following notations for each $n \geq 1$, $\Gamma \in \mathcal{B}(\mathbf{X})$ and $h_n = (y_0, \theta_1, y_1, \ldots, \theta_n, y_n) \in \mathbf{H}_n$:

$$\begin{aligned}\lambda_n^u(\Gamma|h_n, t) &:= \int_{\mathbf{A}^G} \tilde{q}(\Gamma|\overline{x}(y_n), a)\Pi_n(da|h_n, t),\\ \Lambda_n^u(\Gamma|h_n, t) &:= \int_0^t \lambda_n^u(\Gamma|h_n, s)ds.\end{aligned}$$

where and below, we put $q_{\Delta}(a) := 0$ for each $a \in \mathbf{A}^G$. Now, for each $n \geq 1$, we introduce the stochastic kernel G_n^u on $(0, \infty] \times \mathbf{Y}_{\Delta}$ given $\mathbf{H}_n$ as follows. For each $h_n = (y_0, \theta_1, y_1, \ldots, \theta_n, y_n) \in \mathbf{H}_n$,

$$G_n^u(\{+\infty\} \times \{\Delta\}|h_n) := \delta_{y_n}(\{\Delta\}) + \delta_{y_n}(\mathbf{Y})e^{-\Lambda_n^u(\mathbf{S}|h_n, +\infty)}\Phi_n(\{+\infty\}|h_n), \quad (4.3)$$

and

$$\begin{aligned}G_n^u(dt \times dy|h_n) := \delta_{y_n}(\mathbf{Y}) \big\{&\Gamma_n^1(dy|h_n)e^{-\Lambda_n^u(\mathbf{S}|h_n, t)}\Phi_n(dt|h_n)\\ &+ \int_{\mathbf{S}} \Phi_n([t, \infty]|h_n)\Gamma_n^0(dy|h_n, t, x)\lambda_n^u(dx|h_n, t)e^{-\Lambda_n^u(\mathbf{S}|h_n, t)}dt\Big\} \quad (4.4)\end{aligned}$$

on $(0, \infty) \times \mathbf{Y}$. For each fixed initial state $x_0 \in \mathbf{S}$, by the Ionescu-Tulcea theorem, see e.g., Proposition C.10 in [57], there exists a probability $\mathrm{P}_{x_0}^u$ on $(\Omega, \mathcal{F})$ such that the restriction of $\mathrm{P}_{x_0}^u$ to $(\Omega, \mathcal{F}_0)$ is given by

$$\mathrm{P}_{x_0}^u\left(\left(\{y_0\} \times \{0\} \times \Gamma \times ((0, \infty] \times \mathbf{Y}_{\Delta})^{\infty}\right) \bigcap \Omega\right) \;=\; u_0(\Gamma|x_0) \quad (4.5)$$

for each $\Gamma \in \mathcal{B}(\mathbf{Y})$; and for each $n \geq 1$, under $\mathrm{P}_{x_0}^u$, the conditional distribution of (Y_{n+1}, Θ_{n+1}) given $\mathcal{F}_{T_n} := \sigma(H_n)$ is determined by $G_n^u(\cdot|H_n)$ and the conditional survival function of Θ_{n+1} given $\mathcal{F}_{T_n}$ under $\mathrm{P}_{x_0}^u$ is given by $G_n^u([t, +\infty] \times \mathbf{Y}_{\infty}|H_n)$.

The cost associated with an intervention $y = (x_0, b_0, x_1, b_1, \ldots) \in \mathbf{Y}$ is given

by

$$C^I(y) := \sum_{k=0}^{\infty} c^I(x_k, b_k, x_{k+1}).$$

Here, recall that an intervention consists of the current state, the sequence of impulses applied in turn at the same time moment and the associated post-impulse states; and each impulse b applied at state x results in a cost $c^I(x, b, z)$ if it leads to the post-impulse state z. (We accept that $c^I(x, \Delta, \Delta) := 0$ for all $x \in \mathbf{S}_\Delta$.) With this notation, we now recall the performance measure considered in this section:

$$\mathcal{L}(u, x) \;:=\; \mathrm{E}_x^u \left[e^{\sum_{n=1}^{\infty} \left(C^I(Y_n) + \int_{T_n}^{T_{n+1}} \int_{\mathbf{A}^G} c^G(\bar{x}(\xi_s), a) \Pi_n(da | H_n, s - T_n) ds \right)} \right]$$

for each $x \in \mathbf{S}$ and policy $u \in \mathcal{U}$. Here we recall that $T_1 = \Theta_1 \equiv 0$, see (4.2). To illustrate more explicitly how the policy acts on the impulses, consider the example of only one intervention and null gradual cost $c^G(x, a) \equiv 0$. Then we may write

$$\begin{aligned} \mathrm{E}_x^u \left[e^{C^I(Y_1)} \right] &= \int_{\mathbf{Y}} u_0(dx_0 \times db_0 \times dx_1 \times db_1 \times \ldots | x) e^{\sum_{k=0}^{\infty} c^I(x_k, b_k, x_{k+1})} \\ &= \int_{\mathbf{Y}} u_0(dy | x) e^{C^I(y)}. \end{aligned}$$

More generally, one can compute $\mathrm{E}_x^u \left[e^{C^I(Y_{n+1})} \right] = \mathrm{E}_x^u \left[\mathrm{E}_x^u \left[e^{C^I(Y_{n+1})} | H_n \right] \right]$, where $\mathrm{E}_x^u \left[e^{C^I(Y_{n+1})} | H_n \right]$ can be written out as a similar integral to the case of $n = 0$ using the conditional laws (4.3) and (4.4).

Let the value function $\mathcal{L}^*$ be denoted by

$$\mathcal{L}^*(x) := \inf_{u \in \mathcal{U}} \mathcal{L}(x, u)$$

for each $x \in \mathbf{S}$. A policy u^* satisfying $\mathcal{L}(x, u^*) = \mathcal{L}^*(x)$ for all $x \in \mathbf{S}$ is called

optimal for the gradual-impulse control problem:

$$\text{Minimize over } u \in \mathcal{U} : \mathcal{L}(x, u). \tag{4.6}$$

In this section, we will present conditions on the system primitives that guarantee the existence of an optimal policy in a simple form as defined as follows.

Definition 4.4. A policy u is called deterministic stationary if there exist some measurable mappings (φ, ψ, f) on $\mathbf{S}$, where $\varphi(x) \in \{0, \infty\}$ for each $x \in \mathbf{S}$, ψ and f are $\mathbf{A}^I$-valued and $\mathbf{A}^G$-valued, such that $\Phi_n(\{\infty\}|h_n) = 1$, $\Pi_n(da|h_n, t) = \delta_{f(\bar{x}(y_n))}(da)$ for all $t \geq 0$, and $u_n(\cdot|x) = \Gamma_n^0(\cdot|h_n, t, x) = \beta^\pi(\cdot|x)$ for some deterministic stationary strategy π in the intervention DTMDP model defined by $\pi(\{\Delta\}|x_0, b_0, x_1, b_1, \ldots, x_n) = I\{\varphi(x_n) = \infty\}$, and $\pi(db|x_0, b_0, x_1, b_1, \ldots, x_n) = I\{\varphi(x_n) = 0\}\delta_{\psi(x_n)}(db)$.

In the above definition, Γ_n^1 was left arbitrary, because, under such a deterministic stationary policy, a new intervention is always triggered by a natural jump.

4.3 Optimality results

In this section, we present the main optimality results in this paper. In a nutshell, under quite general conditions on the system primitives of the gradual-impulse control problem (4.6), we show that it can be solved via problem (B.1) in Appendix B for a simple DTMDP model, which we refer to as the tilde DTMDP model. In this way, we show that the gradual-impulse control problem (4.6) admits a deterministic stationary optimal policy.

In order to formulate the tilde DTMDP model, we impose the following condition.

Condition 4.1. There exists an $[1, \infty)$-valued continuous function w on $\mathbf{S}$ such that $c^G(x, a) + q_x(a) + 1 \leq w(x)$ for each $(x, a) \in \mathbf{S} \times \mathbf{A}^G$.

If c^G is a continuous function, then the above condition is a consequence of Condition 4.2 and the Berge theorem, see Proposition 7.32 of [9]. Several statements below do not need the bounding function w in Condition 4.1 to be continuous. In this connection, we also mention that a Borel measurable function w satisfying the inequality in Condition 4.1 always exists, see Lemma 1 of [36] and recall (4.1).

Recall that $\tilde{\mathbf{A}} = \mathbf{A}^I \bigcup \mathbf{A}^G$ is the disjoint union of $\mathbf{A}^G$ and $\mathbf{A}^I$. We are now in position to define the tilde DTMDP model in terms of the system primitives of the gradual-impulse control problem (4.6).

Definition 4.5. The tilde DTMDP model is specified by the following four-tuple $\{\mathbf{S}, \tilde{\mathbf{A}}, \tilde{Q}, \tilde{l}\}$, where $\mathbf{S}$ and $\tilde{\mathbf{A}}$ are its state and action spaces, and its transition probability $\tilde{Q}$ on $\mathbf{S}$ given $\mathbf{S} \times \tilde{\mathbf{A}}$ and cost function $\tilde{l}$ are defined by

$$\tilde{Q}(dy|x,a) := \frac{q(\Gamma|x,a)}{w(x)} + \delta_x(dy),\ \tilde{l}(x,a,y) := \ln \frac{w(x)}{w(x) - c^G(x,a)}$$

for all $a \in \mathbf{A}^G$,

$$\tilde{Q}(dy|x,b) := Q(dy|x,b),\ \tilde{l}(x,b,y) := c^I(x,b,y)$$

for all $b \in \mathbf{A}^I$.

For the solvability of problem (B.1) for the tilde DTMDP model, we impose the following compactness-continuity condition.

Condition 4.2. The functions c^I and c^G are lower semicontinuous on $\mathbf{S} \times \mathbf{A}^I \times \mathbf{S}$ and $\mathbf{S} \times \mathbf{A}^G$, respectively; and for each bounded continuous function g on $\mathbf{S}$, $\int_{\mathbf{S}} g(y) Q(dy|x,b)$ and $\int_{\mathbf{S}} g(y) \tilde{q}(dy|x,a)$ are continuous in $(x,b) \in \mathbf{S} \times \mathbf{A}^I$ and $(x,a) \in \mathbf{S} \times \mathbf{A}^G$, respectively. (Recall also that $\mathbf{A}^G$ and $\mathbf{A}^I$ are compact.)

Under Conditions 4.1 and 4.2, one can easily check that the tilde DTMDP model is semicontinuous, so that the value function W^* for problem (B.1) of the tilde DTMDP model is lower semicontinuous, and there exists an optimal

deterministic stationary strategy for it, see Proposition B.1(f). We collect these observations in the next statement for future reference.

Proposition 4.1. Suppose Conditions 4.1 and 4.2 are satisfied. Then the value function W^* of problem (B.1) for the tilde DTMDP model coincides is the minimal $[1,\infty]$-valued lower semicontinuous function satisfying

$$V(x) = \inf_{\tilde{a}\in\tilde{\mathbf{A}}} \left\{ \int_{\mathbf{S}} e^{\tilde{l}(x,\tilde{a},y)} V(y) \tilde{Q}(dy|x,\tilde{a}) \right\}, \ x \in \mathbf{S}, \tag{4.7}$$

and the above relation holds with equality being replaced by "$\geq$", too. A pair of measurable mappings (ψ^*, f^*) from $\mathbf{S}$ to $\mathbf{A}^I$ and $\mathbf{A}^G$, respectively, is a deterministic optimal stationary strategy for problem (B.1) of the tilde DTMDP model if and only if

$$\begin{aligned} & \int_{\mathbf{S}} e^{\tilde{l}(x,\tilde{a}^*,y)} W^*(y) \tilde{Q}(dy|x,\tilde{a}^*) \\ = & \inf_{\tilde{a}\in\tilde{\mathbf{A}}} \left\{ \int_{\mathbf{S}} e^{\tilde{l}(x,\tilde{a},y)} W^*(y) \tilde{Q}(dy|x,\tilde{a}) \right\} \\ = & \int_{\mathbf{S}} e^{\tilde{l}(x,\psi^*(x),y)} W^*(y) \tilde{Q}(dy|x,\psi^*(x)) I\{\tilde{a}^* \in \mathbf{A}^I\} \\ + & \int_{\mathbf{S}} e^{\tilde{l}(x,f^*(x),y)} W^*(y) \tilde{Q}(dy|x,f^*(x)) I\{\tilde{a}^* \in \mathbf{A}^G\}. \end{aligned} \tag{4.8}$$

Such a pair (ψ^*, f^*) of measurable selectors exists.

We introduce the notation to be used in the next statement. Define for each $[1,\infty]$-valued universally measurable function g on $\mathbf{S}$

$$\mathbf{S}^G(g) := \left\{ x \in \mathbf{S} : \infty > g(x) = \inf_{a\in\mathbf{A}^G} \left\{ \int_{\mathbf{S}} g(y) \tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a)) g(x) \right\} \right\} \tag{4.9}$$

$$\mathbf{S}^I(g) := \left\{ x \in \mathbf{S} : g(x) = \inf_{b\in\mathbf{A}^I} \left\{ \int_{\mathbf{S}} g(y) e^{c^I(x,b,y)} Q(dy|x,b) \right\} \right\}$$

Proposition B.1 in the Appendix asserts that W^* is universally measurable so that the integrals $\int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,a)$ and $\int_{\mathbf{S}} W^*(y) e^{c^I(x,b,y)} Q(dy|x,b)$ are well

defined.

Theorem 4.1. Suppose Conditions 4.1 and 4.2 are satisfied. Then the following assertions hold.

(a) The value function W^* of problem (B.1) for the tilde DTMDP model coincides with $\mathcal{L}^*$.

(b) $\mathbf{S} \setminus \mathbf{S}^I(W^*) \subseteq \mathbf{S}^G(W^*)$.

(c) There is a deterministic stationary optimal policy for the gradual-impulse control problem (4.6), which can be obtained as follows. For each pair (ψ^*, f^*) of measurable mappings satisfying (4.8) (and there exists such a pair by Proposition 4.1), the following deterministic stationary policy (φ, ψ, F) is optimal, where

$$\psi(x) = \psi^*(x), \ F(x)_t(da) \equiv \delta_{f^*(x)}(da)$$

for all $x \in \mathbf{S}$, and $\varphi(x) = \infty$ (respectively, $\varphi(x) = 0$) for all $x \in \mathbf{S} \setminus \mathbf{S}^I(W^*)$ (respectively $x \in \mathbf{S}^I(W^*)$).

The proofs and the other statements in this section are postponed to Section 4.5.

According to Theorem 4.1, roughly speaking, if the current state is in $\mathbf{S}^G(W^*)$, then it is optimal not to apply impulse until the next natural jump; and if the current state is in $\mathbf{S}^I(W^*)$, then it is optimal to apply immediately an impulse.

According to Theorem 4.1, (4.7) is the optimality equation for the gradual-impulse control problem (4.6). It can be written out in an equivalent form that does not involve the function w, which might be more convenient sometimes.

Corollary 4.1. Suppose Conditions 4.1 and 4.2 are satisfied. Then the following assertions hold.

(a) $\mathcal{L}^*$ is the minimal $[1,\infty]$-valued lower semicontinuous function on $\mathbf{S}$ satisfying

$$\begin{aligned}&\inf_{a\in\mathbf{A}^G}\left\{\int_{\mathbf{S}}\mathcal{L}^*(y)\tilde{q}(dy|x,a)-(q_x(a)-c^G(x,a))\mathcal{L}^*(x)\right\}\geq 0,\ (4.10)\\&\forall\ x\in\mathbf{S}^*(\mathcal{L}^*):=\{x\in\mathbf{S}:\ \mathcal{L}^*(x)<\infty\}\end{aligned}$$

and

$$\mathcal{L}^*(x)\leq\inf_{b\in\mathbf{A}^I}\left\{\int_{\mathbf{S}}e^{c^I(x,b,y)}\mathcal{L}^*(y)Q(dy|x,b)\right\},\ x\in\mathbf{S},\tag{4.11}$$

whereas at each $x\in\mathbf{S}$, the inequality in either (4.10) or (4.11) holds with equality.

(b) A pair (ψ^*,f^*) of measurable mappings satisfies (4.8) if and only if

$$\begin{aligned}&\inf_{a\in\mathbf{A}^G}\left\{\int_{\mathbf{S}}\mathcal{L}^*(y)\tilde{q}(dy|x,a)-(q_x(a)-c^G(x,a))\mathcal{L}^*(x)\right\}\\=\ &\int_{\mathbf{S}}\mathcal{L}^*(y)\tilde{q}(dy|x,f^*(x))-(q_x(f^*(x))-c^G(x,f^*(x)))\mathcal{L}^*(x)\end{aligned}$$

for each $x\in\mathbf{S}^G(\mathcal{L}^*)$, and

$$\inf_{b\in\mathbf{A}^I}\left\{\int_{\mathbf{S}}e^{c^I(x,b,y)}\mathcal{L}^*(y)Q(dy|x,b)\right\}=\int_{\mathbf{S}}\mathcal{L}^*(y)e^{c^I(x,\psi^*(x),y)}Q(dy|x,\psi^*(x))$$

(According to Theorem 4.1, (ψ^*,f^*) gives rise to a deterministic stationary optimal policy for the gradual-impulse control problem (4.6).)

Under the conditions of the previous statement, in the first glance, given $\mathcal{L}^*$ being an $[1,\infty]$-valued lower semicontinuous function on $\mathbf{S}$, it may be not immediately clear why the claimed measurable selector f^* exists because in

$$\begin{aligned}&\int_{\mathbf{S}}\mathcal{L}^*(y)\tilde{q}(dy|x,a)-(q_x(a)-c^G(x,a))\mathcal{L}^*(x)\\=\ &\left(\int_{\mathbf{S}}\mathcal{L}^*(y)\tilde{q}(dy|x,a)+c^G(x,a)\mathcal{L}^*(x)\right)-(q_x(a)\mathcal{L}^*(x))\end{aligned}$$

the expressions in the two brackets are both lower semicontinuous in $(x, a) \in \mathbf{S} \times \mathbf{A}^G$, and the difference between two lower semicontinuous functions may be not lower semicontinuous. This and Lemma 4.5 are the motivation of considering the tilde DTMDP model.

To end this section, we present a simple example to demonstrate a situation, where it is natural and necessary to allow multiple impulses at a single time moment.

Example 4.2. Let us revisit Example 4.1. The model has a state space $\{1, 2\}$, where 1 stands for the rat being present in the kitchen, and 2 indicates the rat either dead or outside the house. The space of gradual controls is a singleton and will not be indicated explicitly, and the space of impulses is $\mathbf{A}^I = \{0, 1\}$, with 1 or 0 standing for shooting or not. So the inequalities (4.10) and (4.11) for the value function $\mathcal{L}^*$ read:

$$\begin{aligned}
&\mathcal{L}^*(2) = 1;\ \mu\mathcal{L}^*(2) - (\mu - l)\mathcal{L}^*(1) \geq 0\\
&\mathcal{L}^*(1) \leq \min\{e^C p\mathcal{L}^*(2) + e^C(1-p)\mathcal{L}^*(1), \mathcal{L}^*(1)\}.
\end{aligned}$$

Suppose $1-e^C(1-p) > 0$. By Theorem 4.1 and Corollary 4.1, if $\frac{e^C p}{1-e^C(1-p)} > \frac{\mu}{\mu-l} > 0$, then $\mathcal{L}^*(1) = \frac{\mu}{\mu-l}$, and the optimal deterministic stationary policy is to never shoot at the rat; otherwise, $\mathcal{L}^*(1) = \frac{e^C p}{1-e^C(1-p)} = E[e^{CZ}]$ with Z following the geometric distribution with success probability p, and the optimal deterministic stationary policy is to keep shooting as soon as the rat is in kitchen until the rat was hit.

The proofs of the statements in this section are based on the investigation of an optimal control problem for another DTMDP model, which will be referred to as the hat DTMDP model and introduced in Section 4.4. For this moment, we point out that the hat DTMDP model is quite different from the tilde DTMDP model: it is with a more complicated action space, and is not necessarily semicontinuous under Conditions 4.1 and 4.2, see Examples 4.3 and 4.4.

4.4 The hat DTMDP model

In this section, we describe a DTMDP problem, which will serve the investigations of the gradual-impulse control problem. To distinguish it from the intervention DTMDP model, we shall refer to it as the hat DTMDP model. The system primitives of the DTMDP model are defined in terms of those of the gradual-impulse control problem. We will reveal, in greater detail, the connections relevant to this chapter between the hat DTMDP problem and the gradual-impulse control problem at the end of this section. For a first impression, roughly speaking, the state process of the hat DTMDP model comes from the system dynamics of the gradual impulse control problem in the following way. The state has two coordinates. Along the (discrete-time) state process of the hat DTMDP model, the second coordinates record the system states of the graduate-impulse control problem immediately after a natural jump (of the marked point process $\{(T_n, Y_n)\}_{n=1}^{\infty}$) or an "actual" impulse (thus the state immediately after the psuedo impulse Δ will not be recorded). The first coordinates record the time in the gradual-impulse control problem elapsed between two consecutive states as recorded in the second coordinates. For the sake of illustration, the realization of the state process in the hat DTMDP model corresponding to the sample path in Figure 1 of the gradual-impulse control problem is displayed in Figure 2.

The hat DTMDP is with a more complicated action space as compared with the original gradual-impulse control problem by using the relaxed control space $\mathcal{R}$ on $\mathbf{A}^G$.

Below we shall use, without special reference, the following notation. If μ is a measure on a Borel space $(\mathbf{S}, \mathcal{B}(\mathbf{S}))$, then the notation $f(\mu) := \int_{\mathbf{S}} f(x)\mu(dx)$ is in use for each measurable function f on $(\mathbf{S}, \mathcal{B}(\mathbf{S}))$, provided that the integral is well defined.

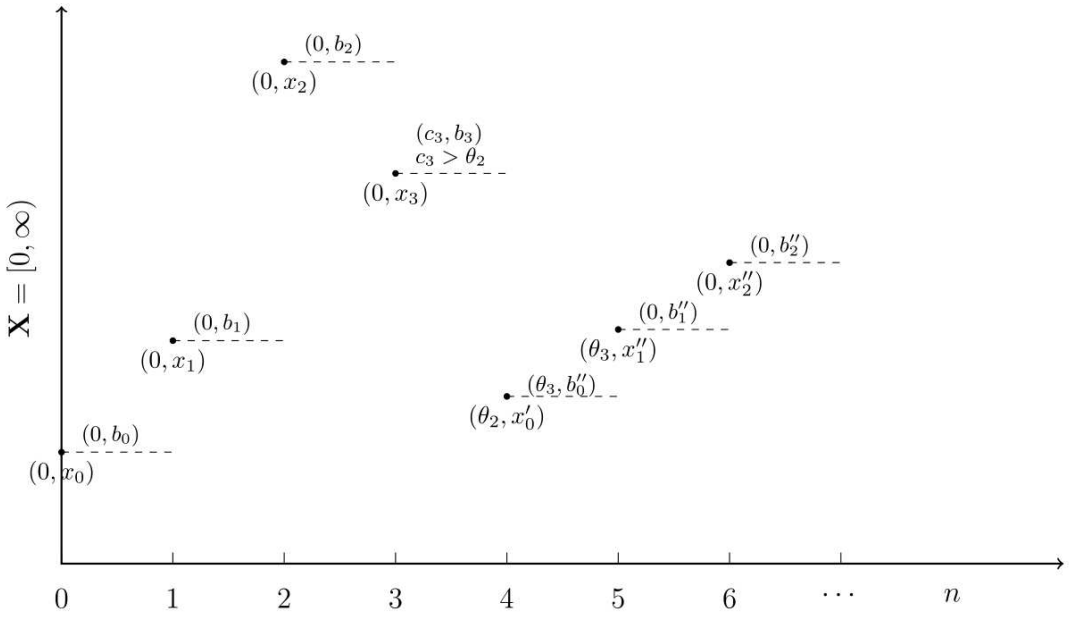

Figure 2: The realization of the state process in the hat DTMDP model corresponding to sample path in the gradual-impulse control in Figure 1. The time index is discrete from $\{0, 1, \dots\}$. The realizations of the components $\{(C_n, B_n)\}_{n=0}^{\infty}$ in the action process $\{\hat{A}_n\}_{n=0}^{\infty}$ are indicated above the dashed lines between consecutive states. For example, $(0, b_0)$ next to the state $(0, x_0)$ indicates that the decision maker applies an impulse b_0 immediately, which results in the next state $(0, x_1)$. All the components $x_0, x_1, \dots, x_0'$, x_1'', x_2'' and $b_1, b_2, b_0'', b_1'', b_2''$ are the same as in Figure 1. The only exception is (c_3, b_3), which does not appear in Figure 1. Nevertheless, $c_3 > \theta_2$, because in Figure 1, the first jump in the marked point process therein at the time moment $\theta_1 + \theta_2 = \theta_2$ is triggered by a natural jump.

4.4.1 Primitives of the hat DTMDP model

The state space of the hat DTMDP model is $\hat{\mathbf{S}} := \{(\infty, x_\infty)\} \bigcup [0,\infty) \times \mathbf{S}$, where (∞, x_∞) is an isolated point, and the action space of the DTMDP is $\hat{\mathbf{A}} := [0,\infty] \times \mathbf{A}^I \times \mathcal{R}$. Endowed with the product topology, where $[0,\infty]$ is compact in the standard topology of the extended real-line, $\hat{\mathbf{A}}$ is also a compact Borel space. Here, $\mathbf{S}$, $\mathbf{A}^I$ and $\mathbf{A}^G$ are the state, impulse and gradual action spaces in the gradual-impulse control problem.

The transition probability p is defined as follows, where the notation introduced above this subsection is in use, e.g., $q_x(\rho_t) := \int_{\mathbf{A}^G} q_x(a)\rho_t(da)$ and $c^G(x,\rho_t) := \int_{\mathbf{A}^G} c^G(x,a)\rho_t(da)$. For each bounded measurable function g on $\hat{\mathbf{S}}$ and action $\hat{a} = (c,b,\rho) \in \hat{\mathbf{A}}$,

$$\begin{aligned}
& \int_{\hat{\mathbf{S}}} g(t,y)p(dt\times dy|(\theta,x),\hat{a}) \\
:= \;& I\{c=\infty\}\left\{ g(\infty,x_\infty)e^{-\int_0^\infty q_x(\rho_s)ds} + \int_0^\infty \int_{\mathbf{S}} g(t,y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t q_x(\rho_s)ds}dt \right\} \\
& + I\{c<\infty\}\left\{ \int_0^c \int_{\mathbf{S}} g(t,y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t q_x(\rho_s)ds}dt + e^{-\int_0^c q_x(\rho_s)ds}\int_{\mathbf{S}} g(c,y)Q(dy|x,b) \right\} \\
= \;& \int_0^c \int_{\mathbf{S}} g(t,y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t q_x(\rho_s)ds}dt + I\{c=\infty\}g(\infty,x_\infty)e^{-\int_0^\infty q_x(\rho_s)ds} \\
& + I\{c<\infty\}e^{-\int_0^c q_x(\rho_s)ds}\int_{\mathbf{S}} g(c,y)Q(dy|x,b)
\end{aligned}$$

for each state $(\theta,x) \in [0,\infty)\times\mathbf{S}$; and

$$\int_{\hat{\mathbf{S}}} g(t,y)p(dt\times dy|(\infty,x_\infty),\hat{a}) := g(\infty,x_\infty).$$

It is known, see e.g., [19, 42], that for each bounded measurable function g on $\hat{\mathbf{S}}$, the above expressions are indeed measurable on $\hat{\mathbf{S}}\times\hat{\mathbf{A}}$, and the same also concerns the cost function l on $\hat{\mathbf{S}}\times\hat{\mathbf{A}}\times\hat{\mathbf{S}}$ defined as follows:

$$l((\theta,x),\hat{a},(t,y)) := I\{(\theta,x)\in[0,\infty)\times\mathbf{S}\}\left\{\int_0^t c^G(x,\rho_s)ds + I\{t=c\}c^I(x,b,y)\right\}$$

for each $(\theta,x),\hat{a},(t,y)) \in \hat{\mathbf{S}}\times\hat{\mathbf{A}}\times\hat{\mathbf{S}}$, accepting that $c^I(x,b,x_\infty)\equiv 0$. Recall that the generic notation $\hat{a} = (c,b,\rho) \in \hat{\mathbf{A}}$ of an action in this hat DTMDP model has

been in use. The pair (c, b) is the pair of the planned time until the next impulse and the next planned impulse, and ρ is (the rule of) the relaxed control to be used during the next sojourn time. The realization of the components $\{(C_n, B_n)\}_{n=0}^{\infty}$ of the action process in the hat DTMDP model corresponding to the sample path in Figure 1 of the gradual-impulse control problem is displayed in Figure 2.

For the convenience in future reference, we make the following definition.

Definition 4.6. The hat DTMDP model is the following four-tuple $\{\hat{\mathbf{S}}, \hat{\mathbf{A}}, p, l\}$, all defined above in terms of the primitives of the gradual-impulse control problem.

Note that Condition 4.2 does not imply that the hat DTMDP model is semi-continuous, which is defined in the appendix. In fact, the transition probability p, in general, does not satisfy the weak continuity condition, even under Condition 4.2. This is demonstrated by the next two examples.

Example 4.3. Suppose $q_x(a) \equiv 0$, and $\mathbf{A}^G$ and $\mathbf{A}^I$ are both singletons. Consider $\hat{a}_n = (c_n, b, \rho)$, where $c_n \to \infty$ and $c_n \in [0, \infty)$ for each $n \geq 1$; and the bounded continuous function on $\hat{\mathbf{S}}$: $g(t, x) \equiv 1$ for each $(t, x) \in [0, \infty) \times \mathbf{S}$, and $g(\infty, x_\infty) = 0$. Then $\int_{\hat{\mathbf{S}}} g(t, y) p(dt \times dy | (\theta, x), \hat{a}_n) = \int_{\mathbf{S}} g(c_n, y) Q(dy|x, b) = 1$ for each $n \geq 1$, whereas $\int_{\hat{\mathbf{S}}} g(t, y) p(dt \times dy | (\theta, x), (\infty, b, \rho)) = g(\infty, x_\infty) = 0 \neq 1$.

Example 4.4. Consider $\mathbf{A}^G = [0, 1]$, $\mathbf{A}^I$ an arbitrary compact Borel space, $\mathbf{S}$ a finite set (endowed with discrete topology), $q_x(a) = a$ for each $x \in \mathbf{S}$. Then consider $x^{(n)} \equiv x \in \mathbf{S}$, $b^{(n)} \equiv b$, $c^{(n)} \equiv c = \infty$, and for each $t \geq 0$, $\rho_t^{(n)}(da) = \delta_{\frac{1}{n}}(da)$, and $\rho_t(da) = \delta_0(da)$. Then for each strongly integrable Caratheodory function $g(t, a)$,

$$\int_0^\infty g(t, \rho_t^{(n)}) dt - \int_0^\infty g(t, \rho_t^{(0)}) dt = \int_0^\infty (g(t, \frac{1}{n}) - g(t, 0)) dt \to 0$$

as $n \to \infty$, by using the dominated convergence theorem. Thus, $\rho^{(n)} \to \rho$ as $n \to \infty$. Let $\hat{a}_n = (c^{(n)}, b^{(n)}, \rho^{(n)})$ and $\hat{a} = (c, b, \rho)$. It follows that $((\theta, x), \hat{a}_n) \to$

$((\theta, x), \hat{a})$ as $n \to \infty$. Now consider the bounded continuous function on $\hat{\mathbf{S}}$ given by $g(\infty, x_\infty) = 1$ and $g(t, x) \equiv 0$ on $[0, \infty) \times \mathbf{S}$. (Recall that (∞, x_∞) is an isolated point in $\hat{\mathbf{S}}$.) Then we see

$$\begin{aligned} & \lim_{n\to\infty} \int_{\hat{\mathbf{S}}} g(t,y) p(dt \times dy | (\theta, x), \hat{a}_n) = \lim_{n\to\infty} e^{-\int_0^\infty q_x(\rho_s^{(n)}) ds} = \lim_{n\to\infty} e^{-\int_0^\infty \frac{1}{n} ds} = 0 \\ \neq \quad & 1 = e^{-\int_0^\infty 0 ds} = \int_{\hat{\mathbf{S}}} g(t,y) p(dt \times dy | (\theta, x), \hat{a}). \end{aligned}$$

Remark 4.1. Example 4.4 implies that the assertion of Lemma 5.12 in [108] (stated without proof) is inaccurate without further conditions (such as $q_x(a) > \delta > 0$ for some $\delta > 0$). Similarly, Lemma 4.1(b) in [46] is correct if $q_x(a) > \delta > 0$ for some $\delta > 0$. However, the optimality results in [108] all survive without assuming extra conditions, as a particular consequence of the arguments presented below in the present chapter.

We use the notation $\hat{h}_n = ((\theta_0, x_0), (c_0, b_0, \rho_0), (\theta_1, x_1), (c_1, b_1, \rho_1) \dots (\theta_n, x_n))$ for the n-history in the hat DTMDP model.

The concerned optimal control problem for the hat DTMDP model reads:

$$\text{Minimize over } \sigma\text{: } \mathbb{E}_{\hat{x}}^{\sigma} \left[e^{\sum_{n=0}^{\infty} l(\hat{S}_n, \hat{A}_n, \hat{S}_{n+1})} \right] =: V((\theta, x), \sigma) \tag{4.12}$$

where $\{\hat{S}_n\}_{n=0}^{\infty}$ and $\{\hat{A}_n\}_{n=0}^{\infty}$ are the state and action processes, and the minimization problem is over all strategies σ in the hat DTMDP model. (See the appendix for the basic notations in a DTMDP.) We denote by V^* the value function of this optimal control problem, i.e.,

$$V^*(\theta, x) := \inf_{\sigma} \mathbb{E}_{\hat{x}}^{\sigma} \left[e^{\sum_{n=0}^{\infty} l(\hat{S}_n, \hat{A}_n, \hat{S}_{n+1})} \right]$$

for each $\hat{x} = (\theta, x) \in \hat{\mathbf{S}}$, where the infimum is over all strategies. Clearly, $V^*(\infty, x_\infty) = 1$. It will be seen in Lemma 4.1 below that V^* depends on (θ, x)

only through x, and a strategy σ is optimal if

$$V((0,x),\sigma) = V^*(x)$$

for each $x \in \mathbf{S}$. Below, when the context is clear, we often consider the restriction of V^* on $\mathbf{S}$ but still use the same notation. The definition of an optimal strategy and other relevant notions of DTMDP are collected in the appendix.

Consider a strategy $\sigma = \{\sigma_n\}_{n=0}^{\infty}$ in the hat DTMDP model, where for each $n \geq 0$, $\sigma_n(d\hat{a}|\hat{h}_n)$ is a stochastic kernel on $\hat{\mathbf{A}}$ given $\hat{h}_n$, which specifies the conditional distribution of the next action (c, b, ρ) given $\hat{h}_n$.

In general, a strategy in the hat DTMDP model can make use of past decision rules of relaxed controls, and the selection of the next relaxed control, and that of the next planned impulse time and impulse do not have to be (conditionally) independent. Therefore, a general strategy in the hat DTMDP model does not immediately correspond to a policy in the continuous-time gradual-impulse control problem described in the previous section. To relate the continuous-time gradual-impulse control problem (4.6) and the hat DTMDP problem (4.12), see Proposition 4.2 below, we introduce the following class of strategies in the hat DTMDP model.

Definition 4.7. A strategy σ in the hat DTMDP model is called typical if under it, given $\hat{h}_n$, the selection of the next action (c, b) and ρ are conditionally independent, and moreover, the selection of ρ is deterministic, i.e.,

$$\sigma_n(dc \times db \times d\rho|\hat{h}_n) = \sigma'_n(dc \times db|\hat{h}_n)\delta_{F^n(\hat{h}_n)}(d\rho),$$

where $F^n(\hat{h}_n)$ is measurable in its argument and takes values in $\mathcal{R}$, and $\sigma'_n(dc \times db|\hat{h}_n)$ is a stochastic kernel on $[0,\infty] \times \mathbf{A}^I$ given $\hat{h}_n$.

One can always write $\sigma'_n(dc \times db|\hat{h}_n) = \varphi_n(dc|\hat{h}_n)\psi_n(db|\hat{h}_n, c)$ for some stochastic kernels φ_n and ψ_n. Intuitively, φ_n defines the (conditional) distribution of the

planned time duration till the next impulse, and $\psi_n(db|\hat{h}_n, c)$ specifies the distribution of the next impulsive action given the history $\hat{h}_n$ and the next impulse moment c, provided that it takes place before the next natural jump. Therefore, we identify a typical strategy $\sigma = \{\sigma_n\}$ as $\{(\varphi_n, \psi_n, F^n)\}_{n=0}^{\infty}$.

For further notational brevity, when the stochastic kernels φ_n are identified with underlying measurable mappings, we will use φ_n for the measurable mappings, and write $\varphi_n(\hat{h}_n)$ instead of $\varphi_n(da|\hat{h}_n)$. The same applies to other stochastic kernels such as ψ_n. The context will exclude any potential confusion.

Finally, in general, we often do not indicate the arguments that do not affect the values of the concerned mappings. For example, if $\varphi_n(\hat{h}_n)$ depends on $\hat{h}_n$ only through x_n, then we write $\varphi_n(da|\hat{h}_n)$ as $\varphi_n(da|x_n)$.

4.4.2 Connection between the gradual-impulse control problem and the hat DTMDP problem

Each policy u as given in Definition 4.3 induces a (typical) strategy $\{(\varphi_n, \psi_n, F^n)\}_{n=0}^{\infty}$ in the hat DTMDP model as follows, where we only need consider $x_n \in \mathbf{S}$, as the definition of the strategies at $x_n = x_\infty$ is immaterial, and can be arbitrary. For each $m \geq 1$, and $h_m \in \mathbf{H}_m$, there exists a strategy $\pi^{\Gamma_m^1, h_m} = \{\pi_n^{\Gamma_m^1, h_m}\}_{n=0}^{\infty}$ in the intervention DTMDP model such that $\Gamma_m^1(dy|h_m) = \beta^{\pi^{\Gamma_m^1, h_m}}(dy|\bar{x}(y_m))$. Similarly, for each $x \in \mathbf{S}, t > 0$, there exists a strategy $\pi^{\Gamma_m^0, h_m, t, x} = \{\pi_n^{\Gamma_m^0, h_m, t, x}\}_{n=0}^{\infty}$ in the intervention DTMDP model such that $\Gamma_m^0(dy|h_m, t, x) = \beta^{\pi^{\Gamma_m^0, h_m, t, x}}(dy|x)$. Finally, there is a strategy $\pi^{u_0} = \{(\pi_n^{u_0})\}_{n=0}^{\infty}$ in the intervention DTMDP model satisfying

$$u_0(dy|x) = \beta^{\pi^{u_0}}(dy|x) \tag{4.13}$$

for each $x \in \mathbf{S}$.

Consider the case of $n = 0$ and let $u_0(\cdot|x) \triangleq \beta^{\pi_{u_0}}(\cdot|x)$ for some strategy

$\pi_{u_0} = \{\pi_n^{u_0}\}_{n=0}^{\infty}$. Then we define

$$\begin{aligned}
\varphi_0(\{0\}|\theta, x) &:= 1 - \pi_0^{u_0}(\{\Delta\}|x); \\
\varphi_0(dc|\theta, x) &:= \pi_0^{u_0}(\{\Delta\}|x)\Phi_1(dc|(x, \Delta, \Delta, \dots), 0, (x, \Delta, \Delta, \dots)) \text{ on } (0, \infty]; \\
\psi_0(db|\theta, x, c) &:= \frac{\pi_0^{u_0}(db|x)}{1 - \pi_0^{u_0}(\{\Delta\}|x)} I\{c = 0\} + I\{c > 0\} \frac{\pi_0^{\Gamma_1^1, ((x,\Delta,\dots),0,(x,\Delta,\dots))}(db|x)}{1 - \pi_0^{\Gamma_1^1, ((x,\Delta,\dots),0,(x,\Delta,\dots))}(\{\Delta\}|x)} \\
&= \frac{\pi_0^{u_0}(db|x)}{1 - \pi_0^{u_0}(\{\Delta\}|x)} I\{c = 0\} + I\{c > 0\} \pi_0^{\Gamma_1^1, ((x,\Delta,\dots),0,(x,\Delta,\dots))}(db|x); \\
F^0(\theta, x)_t(da) &:= \Pi_1(da|(x, \Delta, \Delta, \dots), 0, (x, \Delta, \Delta, \dots), t)
\end{aligned}$$

where the second equality in the definition of $\psi_0(db|\theta, x, c)$ holds because

$$\pi_0^{\Gamma_1^1, ((x,\Delta,\dots),0,(x,\Delta,\dots))}(\{\Delta\}|x) = 0$$

which follows from the requirement that $\Gamma_n^1(\cdot|h_n) \in \mathcal{P}^{\mathbf{Y}^*}(\bar{x}(y_n))$ for all $n \geq 1$ in Definition 4.3. Also concerning the definition of $\psi_0(db|\theta, x, c)$, note that if the denominator in $1 - \pi_0^{u_0}(\{\Delta\}|x) = 0$, we put $\frac{\pi_0^{u_0}(db|x)}{1-\pi_0^{u_0}(\{\Delta\}|x)}$ as an arbitrary stochastic kernel. The reason is that in the expression $\frac{\pi_0^{u_0}(db|x)}{1-\pi_0^{u_0}(\{\Delta\}|x)} I\{c = 0\}$, equality $1 - \pi_0^{u_0}(\{\Delta\}|x) = 0$ would indicate that the probability of selecting an instantaneous impulse is zero, and so $I\{c = 0\} = 0$ almost surely. The same explanation applies to the definitions of $\psi_n(db|\hat{h}_n, c)$ below, and will not be repeated there. Note that the right hand side does not depend on $\theta \in [0, \infty)$, because the initial time moment is always fixed to be $\theta = 0$.

The intuition behind the above definition of (φ_0, ψ_0, F^0) is as follows. Recall that, if the initial system state is $x \in \mathbf{S}$, then the intervention $y_1 \in \mathbf{Y}$ at the initial time in the gradual-impulse control problem is a realization from the distribution $u_0(\cdot|x) = \beta^{\pi^{u_0}}(\cdot|x)$, which is the strategic measure of some strategy $\pi^{u_0} = \{\pi_n^{u_0}\}_{n=0}^{\infty}$ in the intervention DTMDP model, see (4.13). Then $\pi_0^{u_0}(\{\Delta\}|x)$ is the probability that no impulse is applied at the initial time 0 (given the initial system state x) in the gradual-impulse control problem. Consequently, $1 - \pi_0^{u_0}(\{\Delta\}|x)$ is the probability to apply an impulse immediately, i.e., to wait time 0 until the next impulse, and thus $\varphi_0(\{0\}|\theta, x)$. This quantity does not depend on θ, because the initial time is always 0. Then for a measurable subset

$\Gamma_1 \subseteq (0, \infty]$,

$$
\begin{aligned}
& \pi_0^{u_0}(\{\Delta\}|x)\Phi_1(\Gamma_1|(x,\Delta,\Delta,\dots),0,(x,\Delta,\Delta,\dots)) \\
= \;& \text{Probability (no impulse at initial time 0 given initial system state } x) \\
& \times \text{Probability (time to wait until next impulse is in } \Gamma_1 \\
& \text{given no impulse is immediately applied at the initial time with the initial state } x),
\end{aligned}
$$

which is equal to

$$
\begin{aligned}
& \text{Probability (No immediate impulse, time duration until the next planned impulse is in } \Gamma) \\
= \;& \text{Probability (the time duration until the next planned impulse is in } \Gamma),
\end{aligned}
$$

and thus $\varphi_0(\Gamma|\theta, x)$, where the equality follows because $\Gamma \subseteq (0, \infty]$. (Recall that a planned impulse takes place if no natural jump occurs during the time duration to wait for it.) Finally, as for $\psi_0(db|\theta, x, c)$, if $c = 0$, and $\Gamma_2 \in \mathcal{B}(\mathbf{A}^I)$, then

$$
\begin{aligned}
& \frac{\pi_0^{u_0}(\Gamma_2|x)}{1 - \pi_0^{u_0}(\{\Delta\}|x)} = \frac{\text{Probability (an immediate impulse from } \Gamma_2 \text{ is applied)}}{\text{Probability(an immediate impulse is applied)}} \\
= \;& \text{Probability (an impulse is applied immediately from } \Gamma_2 \\
& \text{given that an impulse is applied after time duration 0)},
\end{aligned}
$$

which is thus $\psi_0(\Gamma_2|\theta, x, 0)$. One can understand $\psi_0(db|\theta, x, c)$ when $c > 0$ in the same manner. The very similar intuition guides the definition of (φ_n, ψ_n, F^n) below.

Now consider $n \geq 1$. Let $\hat{h}_n = ((\theta_0, x_0), (c_0, b_0, \rho_0), (\theta_1, x_1), (c_1, b_1, \rho_1) \dots (\theta_n, x_n))$ be the n-history in the hat DTMDP model. If $\{1 \leq i \leq n : \ \theta_i > 0\} = \emptyset$, then

we define

$$\varphi_n(\{0\}|\hat{h}_n) := 1 - \pi_n^{u_0}(\{\Delta\}|x_0, b_0, \dots, b_{n-1}, x_n),$$
$$\varphi_n(dc|\hat{h}_n) := \pi_n^{u_0}(\{\Delta\}|x_0, b_0, \dots, b_{n-1}, x_n)\Phi_1(dc|y_0, 0, (x_1, b_1, \dots, x_n, \Delta, \Delta, \dots)) \text{ on } (0, \infty];$$
$$\psi_n(db|\hat{h}_n, c) := \frac{\pi_n^{u_0}(db|x_0, b_0, x_1, b_1, \dots, x_n)}{1 - \pi_n^{u_0}(\{\Delta\}|x_0, b_0, x_1, b_1, \dots, x_n)} I\{c = 0\}$$
$$+ I\{c > 0\} \frac{\pi_0^{\Gamma_1^1, (y_0, 0, (x_0, b_0, \dots, x_n, \Delta, \dots))}(db|x_n)}{1 - \pi_0^{\Gamma_1^1, (y_0, 0, (x_0, b_0, \dots, x_n, \Delta, \dots))}(\{\Delta\}|x_n)}$$
$$= \frac{\pi_n^{u_0}(db|x_0, b_0, x_1, b_1, \dots, x_n)}{1 - \pi_n^{u_0}(\{\Delta\}|x_0, b_0, x_1, b_1, \dots, x_n)} I\{c = 0\} + I\{c > 0\} \pi_0^{\Gamma_1^1, (y_0, 0, (x_0, b_0, \dots, x_n, \Delta, \dots))}(db|x_n);$$
$$F^n(\hat{h}_n)_t(da) := \Pi_1(da|y_0, 0, (x_0, b_0, \dots, x_n, \Delta, \Delta, \dots), t).$$

Recall that $y_0 = (x_0, \Delta, \Delta, \dots)$.

If $\{1 \le i \le n : \ \theta_i > 0\} \neq \emptyset$, then let $m(\hat{h}_n) := \#\{1 \le i \le n : \ \theta_i > 0\}$, and $l(\hat{h}_n) := \max\{1 \le i \le n : \theta_i > 0\}$. When the context is clear, we write m and l instead of $m(\hat{h}_n)$ and $l(\hat{h}_n)$ for brevity. Let h_m be the m-history in the gradual-impulse control problem contained in $\hat{h}_n$. More precisely, h_m is defined based on $\hat{h}_n$ as follows. Let $\tau_0(\hat{h}_n) = 0$, and $\tau_i(\hat{h}_n) := \inf\{j > \tau_{i-1} : \ \theta_j > 0\}$ for each $i \ge 1$. Note that $l = \tau_m$. Then $h_m = h_m(\hat{h}_n) = (y_0, 0, y_1, \theta_{\tau_1}, y_2, \dots, \theta_{\tau_{m-1}}, y_m)$, where

$$y_0 = (x_0, \Delta, \Delta, \dots); \ y_1 = (x_0, b_0, x_1, b_1, \dots, x_{\tau_1 - 1}, \Delta, \Delta, \dots);$$
$$\text{if } \theta_{\tau_1} = c_{\tau_1 - 1}, \text{ then } y_2 = (x_{\tau_1 - 1}, b_{\tau_1 - 1}, x_{\tau_1}, b_{\tau_1}, \dots, x_{\tau_2 - 1}, \Delta, \Delta, \dots),$$
$$\text{if } \theta_{\tau_1} < c_{\tau_1 - 1}, \text{ then } y_2 = (x_{\tau_1}, b_{\tau_1}, \dots, x_{\tau_2 - 1}, \Delta, \Delta, \dots);$$
$$\vdots$$
$$\text{if } \theta_{\tau_{m-1}} = c_{\tau_{m-1} - 1}, \text{ then } y_m = (x_{\tau_{m-1} - 1}, \dots, x_{\tau_m - 1}, \Delta, \Delta, \dots),$$
$$\text{if } \theta_{\tau_{m-1}} < c_{\tau_{m-1} - 1}, \text{ then } y_m = (x_{\tau_{m-1}}, \dots, x_{\tau_m - 1}, \Delta, \Delta, \dots).$$

For example, if

$$\hat{h}_5 = ((0, x_0), (b_0, 0, \rho^0), (0, x_1), (b_1, 3, \rho^1), (3, x_2), (b_2, 0, \rho^2), (0, x_3), (b_3, 2, \rho^3), (1, x_4), (b_4, 0, \rho^4), (0, x_5)),$$

then $n = 5$, $m = 2$, $l = 4$, $\tau_1 = 2$, $\tau_2 = 4$, and $h_2 = (y_0, 0, y_1, 3, y_2)$ with $y_1 = (x_0, b_0, x_1, \Delta, \dots)$ and $y_2 = (x_1, b_1, x_2, b_2, x_3, \Delta, \dots)$. Roughly speaking, the integer $m(\hat{h}_n)$ counts the number of interventions (except y_0) contained in the n-history of the hat DTMDP model.

If $0 < \theta_l = c_{l-1}$, we define

$$
\begin{aligned}
&\varphi_n(\{0\}|\hat{h}_n) := 1 - \pi^{\Gamma^1_m, h_m}_{n-l+1}(\{\Delta\}|x_{l-1}, b_{l-1}, \dots, b_{n-1}, x_n),\\
&\varphi_n(dc|\hat{h}_n) := \pi^{\Gamma^1_m, h_m}_{n-l+1}(\{\Delta\}|x_{l-1}, b_{l-1}, \dots, b_{n-1}, x_n)\Phi_m(dc|h_m) \text{ on } (0, \infty];\\
&\psi_n(db|\hat{h}_n, c) := \frac{\pi^{\Gamma^1_m, h_m}_{n-l+1}(db|x_{l-1}, b_{l-1}, \dots, b_{n-1}, x_n)}{1 - \pi^{\Gamma^1_m, h_m}_{n-l+1}(\{\Delta\}|x_{l-1}, b_{l-1}, \dots, b_{n-1}, x_n)} I\{c = 0\}\\
&+ I\{c > 0\}\frac{\pi_0^{\Gamma^1_{m+1}, (h_m, \theta_l, (x_{l-1}, b_{l-1}, \dots, x_n, \Delta, \dots))}(db|x_n)}{1 - \pi_0^{\Gamma^1_{m+1}, (h_m, \theta_l, (x_{l-1}, b_{l-1}, \dots, x_n, \Delta, \dots))}(\{\Delta\}|x_n)}\\
&= \frac{\pi^{\Gamma^1_m, h_m}_{n-l+1}(db|x_{l-1}, b_{l-1}, \dots, b_{n-1}, x_n)}{1 - \pi^{\Gamma^1_m, h_m}_{n-l+1}(\{\Delta\}|x_{l-1}, b_{l-1}, \dots, b_{n-1}, x_n)} I\{c = 0\}\\
&+ I\{c > 0\}\pi_0^{\Gamma^1_{m+1}, (h_m, \theta_l, (x_{l-1}, b_{l-1}, \dots, x_n, \Delta, \dots))}(db|x_n);\\
&F^n(\hat{h}_n)_t(da) := \Pi_m(da|h_m, t).
\end{aligned}
$$

Finally, if $0 < \theta_l < c_{l-1}$, then we define

$$
\begin{aligned}
&\varphi_n(\{0\}|\hat{h}_n) := 1 - \pi^{\Gamma^0_m, h_m, \theta_l, x_l}_{n-l}(\{\Delta\}|x_l, b_l, \dots, b_{n-1}, x_n),\\
&\varphi_n(dc|\hat{h}_n) := \pi^{\Gamma^0_m, h_m, \theta_l, x_l}_{n-l}(\{\Delta\}|x_l, b_l, \dots, b_{n-1}, x_n)\Phi_m(dc|h_m) \text{ on } (0, \infty];\\
&\psi_n(db|\hat{h}_n, c) := \frac{\pi^{\Gamma^0_m, h_m, \theta_l, x_l}_{n-l}(db|x_l, b_l, \dots, b_{n-1}, x_n)}{1 - \pi^{\Gamma^0_m, h_m, \theta_l, x_l}_{n-l}(\{\Delta\}|x_l, b_l, \dots, b_{n-1}, x_n)} I\{c = 0\}\\
&+ I\{c > 0\}\frac{\pi_0^{\Gamma^1_{m+1}, (h_m, \theta_l, (x_l, b_l, \dots, x_n, \Delta, \dots))}(db|x_n)}{1 - \pi_0^{\Gamma^1_{m+1}, (h_m, \theta_l, (x_l, b_l, \dots, x_n, \Delta, \dots))}(\{\Delta\}|x_n)}\\
&= \frac{\pi^{\Gamma^0_m, h_m, \theta_l, x_l}_{n-l}(db|x_l, b_l, \dots, b_{n-1}, x_n)}{1 - \pi^{\Gamma^0_m, h_m, \theta_l, x_l}_{n-l}(\{\Delta\}|x_l, b_l, \dots, b_{n-1}, x_n)} I\{c = 0\}\\
&+ I\{c > 0\}\pi_0^{\Gamma^1_{m+1}, (h_m, \theta_l, (x_l, b_l, \dots, x_n, \Delta, \dots))}(db|x_n);\\
&F^n(\hat{h}_n)_t(da) := \Pi_m(da|h_m, t).
\end{aligned}
$$

To be specific, we call the (typical) strategy $\sigma = \{(\varphi_n, \psi_n, F^n)\}_{n=0}^{\infty}$ defined above as the strategy induced by the policy u. The next statement reveals a connection between a policy u and its induced strategy σ for the hat DTMDP model.

Proposition 4.2. For each policy u and the strategy $\sigma = \{(\varphi_n, \psi_n, F^n)\}_{n=0}^{\infty}$ induced by u, $\mathcal{L}(x,u) = V((0,x),\sigma)$, and therefore, $\mathcal{L}^*(x) \geq V^*(x)$ for each $x \in \mathbf{S}$.

Proof. One can verify

$$\begin{aligned} & \mathrm{E}_x^u \left[e^{\sum_{i=1}^n C^I(Y_i) + \sum_{i=2}^n \int_0^{\Theta_i} \int_{\mathbf{A}G} c^G(\overline{x}(Y_{i-1}),a)\Pi_{i-1}(da|H_{i-1},s)ds} \right] \\ = \; & \mathbb{E}_{(0,x)}^{\sigma} \left[e^{\sum_{i=0}^{\tau_n-1} c^I(X_i,B_i,X_{i+1}) + \sum_{i=2}^n \int_0^{\Theta_{\tau_{i-1}}} \int_{\mathbf{A}G} c^G(X_{\tau_{i-1}-1},a)F^{\tau_{i-1}-1}(\hat{H}_{\tau_{i-1}-1})_s(da)ds} \right] \end{aligned}$$

for each $n \geq 1$. The case of $n = 1$ can be readily seen (we accept $\sum_{n=2}^{1}(\cdot) := 0$), as a consequence of the definitions of the strategy $\sigma = \{(\varphi_n, \psi_n, F^n)\}_{n=0}^{\infty}$ induced by u. The general case follows from an inductive argument. The cumbersome details are omitted. Passing to the limit as $n \to \infty$ and an application of the monotone convergence theorem yield the equality in the statement. The last assertion holds automatically from the first assertion. □

Remark 4.2. A deterministic stationary policy say u^D is associated with a strategy $\sigma^D = (\varphi, \psi, F)$ in the hat DTMDP model, where $F(x)_t(da) = \delta_{f(x)}(da)$ for all $t \geq 0$. It is evident that $\mathcal{L}(x, u^D) = V(x, \sigma^D)$ for each $x \in \mathbf{S}$. Thus, if the hat DTMDP problem (4.12) has an optimal strategy in the form of $\sigma^D = (\varphi, \psi, F)$, then the previous discussions lead to $\mathcal{L}^*(x) = V^*(x)$, and that the deterministic stationary policy u^D associated with σ^D is optimal for the gradual-impulse control problem (4.6).

4.5 Proof of the main statements

In this section, we prove the results stated in Section 4.3. This is based on the investigation of problem (4.12) for the hat DTMDP model described in Section 4.4. In this section, unless specified otherwise, V^* is understood as the value function of problem (4.12) for the hat DTMDP model. More exactly, the main properties concerning V^* are summarized in the next statement.

Proposition 4.3. (a) V^* is a $[1,\infty]$-valued lower semianalytic function on $\mathbf{S}$ satisfying

$$\inf_{a\in\mathbf{A}^G}\left\{\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))V^*(x)\right\} \geq 0, \quad (4.14)$$
$$\forall\ x \in \mathbf{S}^*(V^*) := \{x \in \mathbf{S} :\ V^*(x) < \infty\}$$

and

$$V^*(x) \leq \inf_{b\in\mathbf{A}^I}\left\{\int_{\mathbf{S}} e^{c^I(x,b,y)}V^*(y)Q(dy|x,b)\right\},\ x \in \mathbf{S}, \quad (4.15)$$

whereas at each $x \in \mathbf{S}$, the inequality in either (4.14) or (4.15) holds with equality.

(b) $\mathbf{S} \setminus \mathbf{S}^I \subseteq \mathbf{S}^G$, where $\mathbf{S}^G := \mathbf{S}^G(V^*)$, see (4.9), and $\mathbf{S}^I := \mathbf{S}^I(V^*)$. (Lemma 4.1 below asserts that V^* is universally measurable so that the integrals $\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a)$ and $\int_{\mathbf{S}} V^*(y)e^{c^I(x,b,y)}Q(dy|x,b)$ are defined.)

Proof. See Lemmas 4.1, 4.3 and 4.4 below. □

Lemma 4.1. The following assertions hold.

(a) The value function V^* depends on the state (θ, x) only through the second coordinate x, and thus we write $V^*(x)$ instead of $V^*(\theta, x)$. The function V^*

is an $[1,\infty]$-valued lower semianalytic function satisfying

$$\begin{aligned} V(x) &= \inf_{\hat{a}\in\hat{\mathbf{A}}} \Bigg\{ \int_0^c \int_{\mathbf{S}} V(y)\tilde{q}(dy|x,\rho_t) e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds} dt \\ &\quad + I\{c=\infty\} e^{-\int_0^\infty q_x(\rho_s)ds} e^{\int_0^\infty c^G(x,\rho_s)ds} \\ &\quad + I\{c<\infty\} e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds} \int_{\mathbf{S}} V(y) e^{c^I(x,b,y)} Q(dy|x,b) \Bigg\}; \\ V(x_\infty) &= 1, \end{aligned} \tag{4.16}$$

and is the minimal $[1,\infty]$-valued lower semianalytic function satisfying the following inequality

$$\begin{aligned} V(x) &\geq \inf_{\hat{a}\in\hat{\mathbf{A}}} \Bigg\{ \int_0^c \int_{\mathbf{S}} V(y)\tilde{q}(dy|x,\rho_t) e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds} dt \\ &\quad + I\{c=\infty\} e^{-\int_0^\infty q_x(\rho_s)ds} e^{\int_0^\infty c^G(x,\rho_s)ds} \\ &\quad + I\{c<\infty\} e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds} \int_{\mathbf{S}} V(y) e^{c^I(x,b,y)} Q(dy|x,b) \Bigg\}; \\ V(x_\infty) &= 1. \end{aligned} \tag{4.17}$$

(b) For each $\epsilon > 0$, there exists an ϵ-optimal deterministic Markov universally measurable strategy that depends on the state (θ, x) only through the second coordinate for the hat DTMDP problem (4.12). (The meaning of universally measurable strategies can be found in Appendix B.)

(c) A deterministic stationary strategy that depends on the state (θ, x) only through x is optimal if and only if it attains the infimum in (4.16) with V^* replacing V, for each $x \in \mathbf{S}$.

(d) For each $x \in \mathbf{S}$, $V^*(x) = \inf_{\pi\in\Pi^U} V(x,\pi)$, where Π^U indicates the class of universally measurable strategies in the hat DTMDP model.

Proof. The fact that the value function V^* is the minimal $[1,\infty]$-valued lower

semianalytic function satisfying

$$\begin{aligned} g(\theta,x) &\geq \inf_{\hat{a}\in\hat{\mathbf{A}}}\Big\{\int_0^c\int_{\mathbf{S}} g(t,y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t(q_x(\rho_s)-c^G(x,\rho_s))ds}dt \\ &\quad +I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds} \\ &\quad +I\{c<\infty\}e^{-\int_0^c(q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\mathbf{S}} g(c,y)e^{c^I(x,b,y)}Q(dy|x,b)\Big\}; \\ g(\infty,x_\infty) &:= 1, \end{aligned}$$

where the inequality can be replaced by equality, follows from Proposition B.1. The existence of an ϵ-optimal deterministic Markov universally measurable strategy follows from Proposition B.1, too. Furthermore, note that the first coordinate in the state (θ,x) does not affect the cost function or the transition probability, from which the independence on the first coordinate of the state (θ,x) follows, c.f. [34]. Now assertions (a,b) follow. Finally, the last two assertions follow from Proposition B.1. □

Lemma 4.2. The function

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

is increasing, for each $x\in\mathbf{S}$ and $\rho\in\mathcal{R}$.

Proof. Let $0\leq t_1<t_2<\infty$ and $x\in\mathbf{S}$ be fixed, and we will verify

$$\begin{aligned} &\int_0^{t_2} e^{-\int_0^\tau(q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau+e^{-\int_0^{t_2}(q_x(\rho_s)-c^G(x,\rho_s))ds}V^*(x) \\ \geq\ &\int_0^{t_1} e^{-\int_0^\tau(q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau+e^{-\int_0^{t_1}(q_x(\rho_s)-c^G(x,\rho_s))ds}V^*(x), \end{aligned}$$

as follows. It is sufficient to consider the case when the left hand side is finite, for otherwise, the above inequality would hold automatically. Then the goal is to

show, by subtracting the right hand side from the left hand side,

$$0 \leq \int_{t_1}^{t_2} e^{-\int_0^\tau (q_x(\rho_s)-c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^{t_2}(q_x(\rho_s)-c^G(x,\rho_s))ds}V^*(x)$$
$$-e^{-\int_0^{t_1}(q_x(\rho_s)-c^G(x,\rho_s))ds}V^*(x).$$

The right hand side of this inequality can be further written as

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

Introduce $\tilde{\rho}_s := \rho_{t_1+s}$ for each $s \geq 0$. The target becomes to show

$$\int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\tilde{\rho}_s)-c^G(x,\tilde{\rho}_s))ds}\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\tilde{\rho}_\tau)d\tau + e^{-\int_0^{t_2-t_1}(q_x(\tilde{\rho}_s)-c^G(x,\tilde{\rho}_s))ds}V^*(x) \geq V^*(x).$$

To this end, for a fixed $\epsilon > 0$, let us consider a deterministic Markov ϵ-optimal universally measurable strategy $\{(\varphi_n^*, \psi_n^*, F^{*,n})\}_{n=0}^\infty$ coming from Lemma 4.1, and an associated universally measurable strategy $\pi^{New} = \{(\varphi_n, \psi_n, F^n)\}_{n=0}^\infty$ defined by $\varphi_0(\theta,x) := \varphi_0^*(x) + t_2 - t_1$, $\psi_0(\theta,x) = \psi_0^*(x)$, $F^0(\theta,x)_s = \tilde{\rho}_s$ if $s \leq t_2 - t_1$ and $F^0(\theta,x)_s = F^{*,0}(\theta,x)_{s-(t_2-t_1)}$ if $s > t_2 - t_1$; and for $n \geq 1$, $\varphi_n((\theta,x),\hat{a},(t,y)) = \varphi_{n-1}^*(y)$, $\psi_n((\theta,x),\hat{a},(t,y)) = \psi_{n-1}^*(y)$, and $F^n((\theta,x),\hat{a},(t,y))_s = F^{*,n-1}(y)_s$ for all $s \geq 0$. Under the universally measurable strategy π^{New}, only the gradual control action $\tilde{\rho}$ is used up to either $t_2 - t_1$ or the natural jump moment, whichever takes place first, after when, the ϵ-optimal universally measurable strategy is in

use, and so

$$
\begin{aligned}
& V^*(x) \le V(x,\pi^{New}) \\
\le\ & \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds} \int_{\mathbf{S}} (V^*(y)+\epsilon)\tilde q(dy|x,\tilde\rho_\tau)d\tau \\
+\ & e^{-\int_0^{t_2-t_1}(q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds}(V^*(x)+\epsilon) \\
=\ & \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde q(dy|x,\tilde\rho_\tau)d\tau + e^{-\int_0^{t_2-t_1}(q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds}V^*(x) \\
& +\epsilon\left(\int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds} q_x(\tilde\rho_\tau)d\tau + e^{-\int_0^{t_2-t_1}(q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds}\right),
\end{aligned}
$$

where the first inequality holds because of the last assertion of Lemma 4.1. Since the expression in the last bracket is nonnegative and finite, and $\epsilon > 0$ was arbitrarily fixed, we see that $V^*(x) \le \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde q(dy|x,\tilde\rho_\tau)d\tau + e^{-\int_0^{t_2-t_1}(q_x(\tilde\rho_s)-c^G(x,\tilde\rho_s))ds}V^*(x)$, as desired. □

Lemma 4.3. Relations (4.14) and (4.15) hold. (Recall from Lemma 4.1 that V^* is universally measurable.)

Proof. Let $x \in \mathbf{S}$ be fixed. Inequality (4.15) immediately follows from Lemma 4.1, if on the right hand side of (4.16) with V^* replacing V, one takes the infimum over actions $\hat a \in \hat{\mathbf{A}}$ with $c = 0$. (Recall the notation in use: $\hat a = (c, b, \rho) \in \hat{\mathbf{A}}$.) Let us verify (4.14) as follows. Suppose $V^*(x) < \infty$. Let $a \in \mathbf{A}^G$ be arbitrarily fixed. If $\int_{\mathbf{S}} V^*(y)\tilde q(dy|x,a) = \infty$, then trivially, $\int_{\mathbf{S}} V^*(y)\tilde q(dy|x,a) - (q_x(a) - c^G(x,a))V^*(x) \ge 0$. Consider the case when $\int_{\mathbf{S}} V^*(y)\tilde q(dy|x,a) < \infty$. Let $t > 0$ be arbitrarily fixed. Then $\int_0^t e^{-\tau(q_x(a)-c^G(x,a))} \int_{\mathbf{S}} V^*(y)\tilde q(dy|x,a)d\tau + e^{-t(q_x(a)-c^G(x,a))}V^*(x)$ is finite. Upon differentiating it with respect to t and applying the fundamental theorem of calculus, we see

$$
e^{-(q_x(a)-c^G(x,a))t} \int_{\mathbf{S}} V^*(y)\tilde q(dy|x,a) - (q_x(a)-c^G(x,a))e^{-t(q_x(a)-c^G(x,a))}V^*(x) \ge 0,
$$

where the inequality follows from Lemma 4.2. Thus, $\int_{\mathbf{S}} V^*(y)\tilde q(dy|x,a) - (q_x(a) - c^G(x,a))V^*(x) \ge 0$. Since $a \in \mathbf{A}^G$ was arbitrarily fixed, we see that (4.14) holds. □

Lemma 4.4. For each $x \in \mathbf{S}$, the inequality in either (4.14) or (4.15) holds with equality.

Proof. Let $x \in \mathbf{S}$ be fixed. If the equality in (4.15) holds at this point, then there is nothing to prove. Suppose the strict inequality holds in (4.15). Then necessarily $V^*(x) < \infty$. The objective is to show that, in this case, (4.14) holds with equality. For the infimum in (4.16) with V^* replacing V, it suffices to consider $c > 0$, because (4.15) holds with strict inequality at the fixed point $x \in \mathbf{S}$ here. Let $\epsilon > 0$ be fixed, and $(c^*, b^*, \rho^*) \in \hat{\mathbf{A}}$ be such that

$$
\begin{aligned}
& V^*(x) + \epsilon \\
\geq\ & \left\{ \int_0^{c^*} \int_{\mathbf{S}} V^*(y) \tilde{q}(dy|x, \rho_t^*) e^{-\int_0^t (q_x(\rho_s^*) - c^G(x, \rho_s^*))ds} dt + I\{c^* = \infty\} e^{-\int_0^\infty q_x(\rho_s^*)ds} \right. \\
& \left. e^{\int_0^\infty c^G(x,\rho_s^*)ds} + I\{c^* < \infty\} e^{-\int_0^{c^*} (q_x(\rho_s^*) - c^G(x,\rho_s^*))ds} \int_{\mathbf{S}} V^*(y) e^{c^I(x,b^*,y)} Q(dy|x,b^*) \right\}
\end{aligned}
$$

There are two cases to be considered: (a) $0 < c^* < \infty$; (b) $c^* = \infty$.

Consider case (a). Then

$$
\begin{aligned}
\epsilon + V^*(x) &\geq \int_0^{c^*} \int_{\mathbf{S}} V^*(y) \tilde{q}(dy|x, \rho_t^*) e^{-\int_0^t (q_x(\rho_s^*) - c^G(x,\rho_s^*))ds} dt \\
&\quad + e^{-\int_0^{c^*} (q_x(\rho_s^*) - c^G(x,\rho_s^*))ds} \int_{\mathbf{S}} V^*(y) e^{c^I(x,b^*,y)} Q(dy|x,b^*) \\
&\geq \inf_{\rho \in \mathcal{R}} \left\{ \int_0^{c^*} e^{-\int_0^t (q_x(\rho_s) - c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y) \tilde{q}(dy|x,\rho_t) dt \right. \\
&+ \left. e^{-\int_0^{c^*} (q_x(\rho_s) - c^G(x,\rho_s))ds} V^*(x) \right\} \\
&\geq V^*(x),
\end{aligned}
$$

where the second inequality holds because of (4.15), and the last inequality holds because of Lemma 4.2. Thus, as $\epsilon > 0$ was arbitrarily fixed,

$$
V^*(x) = \inf_{\rho \in \mathcal{R}} \left\{ \int_0^{c^*} e^{-\int_0^t (q_x(\rho_s) - c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y) \tilde{q}(dy|x,\rho_t) dt + e^{-\int_0^{c^*} (q_x(\rho_s) - c^G(x,\rho_s))ds} V^*(x) \right\} \quad (4.18)
$$

Let $\delta > 0$ be fixed. There is some $\rho \in \mathcal{R}$ such that

$$\int_0^{c^*} e^{-\int_0^t (q_x(\rho_s) - c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_t)dt < \infty, \quad \int_0^{c^*} (q_x(\rho_s) - c^G(x,\rho_s))ds < \infty$$

(for the infimum in (4.18), it suffices to concentrate on such elements of $\mathcal{R}$ as $V^*(x) < \infty$), and

$$\begin{aligned}
\delta &\geq \int_0^{c^*} e^{-\int_0^s (q_x(\rho_v) - c^G(x,\rho_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_s)ds + e^{-\int_0^t (q_x(\rho_s) - c^G(x,\rho_s))ds} V^*(x) - V^*(x) \\
&= \int_0^{c^*} e^{-\int_0^s (q_x(\rho_v) - c^G(x,\rho_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_s)ds \\
&\quad - \int_0^{c^*} (q_x(\rho_\tau) - c^G(x,\rho_\tau)) e^{-\int_0^\tau (q_x(\rho_s) - c^G(x,\rho_s))ds} d\tau V^*(x) \\
&= \int_0^{c^*} e^{-\int_0^s (q_x(\rho_v) - c^G(x,\rho_v))dv} \left\{ \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_s) - (q_x(\rho_s) - c^G(x,\rho_s))V^*(x) \right\} ds \\
&\geq \int_0^{c^*} e^{-\int_0^s (q_x(\rho_v) - c^G(x,\rho_v))dv} ds \inf_{a \in \mathbf{A}^G} \left\{ \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))V^*(x) \right\} \\
&\geq \int_0^{c^*} e^{-\overline{q}_x s} ds \inf_{a \in \mathbf{A}^G} \left\{ \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))V^*(x) \right\} \geq 0,
\end{aligned}$$

where the last inequality holds because of (4.14). Since $\int_0^{c^*} e^{-\overline{q}_x s} ds > 0$ and $\delta > 0$ was arbitrarily fixed, we see that (4.14) holds with equality.

Now consider case (b). Then

$$\epsilon + V^*(x) \geq \inf_{\rho \in \mathcal{R}} \left\{ \int_0^\infty e^{-\int_0^t (q_x(\rho_s) - c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_t)dt + e^{-\int_0^\infty q_x(\rho_s)ds} e^{\int_0^\infty c^G(x,\rho_s)ds} \right\}.$$

One can apply the proof of Lemma 5.3 of [108] to show that for each $t \in [0,\infty)$,

$$V^*(x) = \inf_{\rho \in \mathcal{R}} \left\{ \int_0^t e^{-\int_0^t (q_x(\rho_s) - c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_t)dt + e^{-\int_0^t (q_x(\rho_s) - c^G(x,\rho_s))ds} V^*(x) \right\} \quad (4.19)$$

To improve the readability, we provide the detailed justification of this fact as follows. We only need consider when $t > 0$; the case of $t = 0$ is trivial. Let $\delta > 0$ be arbitrarily fixed. Then there is some $\hat{\rho} \in \mathcal{R}$ such that

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

Define $\tilde{\rho} \in \mathcal{R}$ by $\tilde{\rho}_s = \hat{\rho}_{t+s}$ for each $s \geq 0$. Then, for each $t \geq 0$,

$$
\begin{aligned}
& \epsilon + V^*(x) + \delta \\
\geq\ & \int_0^t e^{-\int_0^\tau (q_x(\hat{\rho}_s) - c^G(x,\hat{\rho}_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau \\
& + \int_t^\infty e^{-\int_0^\tau (q_x(\hat{\rho}_s) - c^G(x,\hat{\rho}_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau \\
& + e^{-\int_0^t (q_x(\hat{\rho}_s) - c^G(x,\hat{\rho}_s))ds} e^{-\int_t^\infty q_x(\hat{\rho}_s)ds} e^{\int_t^\infty c^G(x,\hat{\rho}_s)ds} \\
=\ & \int_0^t e^{-\int_0^\tau (q_x(\hat{\rho}_s) - c^G(x,\hat{\rho}_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau + e^{-\int_0^t (q_x(\hat{\rho}_v) - c^G(x,\hat{\rho}_v))dv} \\
& \times \left\{ \int_0^\infty e^{-\int_0^s (q_x(\tilde{\rho}_v) - c^G(x,\tilde{\rho}_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\tilde{\rho}_s)ds + e^{-\int_0^\infty q_x(\tilde{\rho}_s)ds} e^{\int_0^\infty c^G(x,\tilde{\rho}_s)ds} \right\} \\
\geq\ & \int_0^t e^{-\int_0^\tau (q_x(\hat{\rho}_s) - c^G(x,\hat{\rho}_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\hat{\rho}_\tau)d\tau + e^{-\int_0^t (q_x(\hat{\rho}_v) - c^G(x,\hat{\rho}_v))dv} V^*(x) \\
\geq\ & \inf_{\rho \in \mathcal{R}} \left\{ \int_0^t e^{-\int_0^\tau (q_x(\rho_s) - c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_\tau)d\tau + e^{-\int_0^t (q_x(\rho_v) - c^G(x,\rho_v))dv} V^*(x) \right\} \\
\geq\ & V^*(x),
\end{aligned}
$$

where the second inequality is by Lemma 4.1(a), which in particular, asserts that V^* satisfies (4.16), and the last inequality is by Lemma 4.2. Since $\epsilon > 0$ and $\delta > 0$ were arbitrarily fixed, the above implies (4.19). Comparing (4.19) with (4.18), we see that case (b) is reduced to case (a). □

Lemma 4.5. Let w be a measurable $[1, \infty)$-valued function satisfying the inequality in Condition 4.1, whose existence is guaranteed as mentioned in the paragraph below Condition 4.1. Consider the transition probability $\tilde{p}(dy|x,a)$ on $\mathcal{B}(\mathbf{S})$ given $(x,a) \in \mathbf{S} \times \mathbf{A}^G$ defined by

$$
\tilde{p}(\Gamma|x,a) := \frac{q(\Gamma|x,a)}{w(x)} + \delta_x(dy),\ \forall\ \Gamma \in \mathcal{B}(\mathbf{S}),\ (x,a) \in \mathbf{S} \times \mathbf{A}^G.
$$

Then a $[1, \infty]$-valued lower semianalytic function V^* (here the notation V^* does not necessarily mean the value function) satisfies (4.14) and (4.15), and for each $x \in \mathbf{S}$, either (4.14) or (4.15) holds with equality, if and only if this $[1, \infty]$-valued

lower semianalytic function satisfies (4.15), for each $x \in \mathbf{S}$

$$V^*(x) \leq \inf_{a \in \mathbf{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x,a)} \int_{\mathbf{S}} V^*(y) \tilde{p}(dy|x,a) \right\}, \tag{4.20}$$

and either (4.15) or (4.20) holds with equality, i.e.,

$$V^*(x) = \min \left\{ \inf_{a \in \mathbf{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x,a)} \int_{\mathbf{S}} V^*(y) \tilde{p}(dy|x,a) \right\}, \inf_{b \in \mathbf{A}^I} \left\{ \int_{\mathbf{S}} V^*(y) e^{c^I(x,b,y)} Q(dy|x,b) \right\} \right\} \tag{4.21}$$

Note that (4.20) automatically holds with equality at $x \in \mathbf{S} \setminus \mathbf{S}^*(V^*) := \{x \in \mathbf{S} : \ V^*(x) = \infty\}$. Also note that the function w in the previous lemma does not need be continuous.

Proof. "Only if" part. Consider a $[1, \infty]$-valued lower semianalytic function V^* satisfying (4.14) and (4.15), and for each $x \in \mathbf{S}$, either (4.14) or (4.15) holds with equality. For $x \in \mathbf{S}^*(V^*) = \{x \in \mathbf{S} : \ V^*(x) < \infty\}$, (4.14) implies for each $a \in \mathbf{A}^G$ that $0 \leq c^G(x,a)V^*(x) + \int_{\mathbf{S}} V^*(y) q(dy|x,a) = (c^G(x,a) - w(x))V^*(x) + w(x) \int_{\mathbf{S}} V^*(y) \tilde{p}(dy|x,a)$, and thus

$$V^*(x) \leq \inf_{a \in \mathbf{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x,a)} \int_{\mathbf{S}} V^*(y) \tilde{p}(dy|x,a) \right\},$$

i.e., (4.20) holds. Let $x \in \mathbf{S}^*(V^*)$ be a point where (4.14) holds with equality. Let us verify at this point $x \in \mathbf{S}^*(V^*)$, (4.20) also holds with equality. For each $\epsilon > 0$, there is some $a_\epsilon \in \mathbf{A}^G$ such that $\epsilon \geq c^G(x,a_\epsilon)V^*(x) + \int_{\mathbf{S}} V^*(y) q(dy|x,a_\epsilon)$ so that

$$\begin{aligned}
V^*(x) + \epsilon &\geq V^*(x) + \frac{\epsilon}{w(x) - c^G(x,a_\epsilon)} \\
&\geq V^*(x) + \frac{c^G(x,a_\epsilon)V^*(x) + \int_{\mathbf{S}} V^*(y) q(dy|x,a_\epsilon)}{w(x) - c^G(x,a_\epsilon)} \\
&= \frac{w(x)}{w(x) - c^G(x,a_\epsilon)} \int_{\mathbf{S}} \tilde{p}(dy|x,a_\epsilon) V^*(y) \\
&\geq \inf_{a \in \mathbf{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x,a)} \int_{\mathbf{S}} V^*(y) \tilde{p}(dy|x,a) \right\},
\end{aligned}$$

and thus $V^*(x) \geq \inf_{a\in\mathbf{A}^G}\left\{\frac{w(x)}{w(x)-c^G(x,a)}\int_{\mathbf{S}} V^*(y)\tilde{p}(dy|x,a)\right\}$. The opposite direction of this inequality was seen earlier, and so (4.20) holds with equality at this point. This completes the "Only if" part. The argument for the "If" part is the same, and omitted. □

Remark 4.3. Consider the function V^* in the previous statement. By inspecting the above proof we see the following useful fact: a pair of measurable mappings ψ^* and f^* from $\mathbf{S}$ to $\mathbf{A}^I$ and $\mathbf{A}^G$ satisfy

$$\frac{w(x)}{w(x)-c^G(x,f^*(x))}\int_{\mathbf{S}} V^*(y)\tilde{p}(dy|x,f^*(x)) = \inf_{a\in\mathbf{A}^G}\left\{\frac{w(x)}{w(x)-c^G(x,a)}\int_{\mathbf{S}} V^*(y)\tilde{p}(dy|x,a)\right\}$$

for each $x \in \mathbf{S}$, at which (4.20) holds with equality, and

$$\int_{\mathbf{S}} e^{c^I(x,\psi^*(x),y)}V^*(y)Q(dy|x,\psi^*(x)) = \inf_{b\in\mathbf{A}^I}\left\{\int_{\mathbf{S}} e^{c^I(x,b,y)}V^*(y)Q(dy|x,b)\right\},\ \forall\ x\in\mathbf{S},$$

if and only if

$$\begin{aligned}&\inf_{a\in\mathbf{A}^G}\left\{\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a)-c^G(x,a))V^*(x)\right\}\\ =\ &\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,f^*(x)) - (q_x(f^*(x))-c^G(x,f^*(x)))V^*(x)\end{aligned}$$

for each $x \in \mathbf{S}$, at which 0 coincides with the left hand side, and

$$\int_{\mathbf{S}} e^{c^I(x,\psi^*(x),y)}V^*(y)Q(dy|x,\psi^*(x)) = \inf_{b\in\mathbf{A}^I}\left\{\int_{\mathbf{S}} e^{c^I(x,b,y)}V^*(y)Q(dy|x,b)\right\},\ \forall\ x\in\mathbf{S}.$$

Lemma 4.6. Conditions 4.1 and 4.2 are satisfied. Then $W^*(x) = V^*(x)$ for each $x \in \mathbf{S}$.

Proof. According to Proposition B.1(a,b), the value function W^* for the tilde model is the minimal $[1,\infty]$-valued lower semianalytic function satisfying (4.7) as well as the inequality obtained by replacing the equality in (4.7) by "$\geq$". Let us verify that $W^* = V^*$ as follows. According to Lemmas 4.3, 4.4 and 4.5, the value function V^* is a $[1,\infty]$-valued lower semianalytic function satisfying (4.7),

c.f. (4.21). Therefore, $W^* \leq V^*$ pointwise.

For the opposite direction of this inequality, let $x \in \mathbf{S}$ be fixed. It suffices to show that W^* satisfies (4.17) at the point x. Then, since the point $x \in \mathbf{S}$ was arbitrarily fixed, one could apply Lemma 4.1 to obtain $V^* \leq W^*$ pointwise.

Recall that, as observed in the beginning of this proof, W^* satisfies (4.21). By Lemma 4.5, it satisfies (4.14) and (4.15), one of which holds with equality at this point x. If (4.15) holds with equality for W^* at x, then

$$\begin{aligned} W^*(x) &= \inf_{b\in\mathbf{A}^I}\left\{\int_{\mathbf{S}} W^*(y)e^{c^I(x,b,y)}Q(dy|x,b)\right\} \\ \geq \inf_{\hat{a}\in\hat{\mathbf{A}}}&\left\{\int_0^c\int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t(q_x(\rho_s)-c^G(x,\rho_s))ds}dt\right. \\ &+I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds} \\ &\left.+I\{c<\infty\}e^{-\int_0^c(q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\mathbf{S}} W^*(y)e^{c^I(x,b,y)}Q(dy|x,b)\right\}, \end{aligned}$$

and thus (4.17) is satisfied by W^* at x, as required. Now suppose (4.14) holds with equality for W^* at x. It suffices to consider $W^*(x) < \infty$, for otherwise, (4.17) automatically holds for W^* at x. According to Remark 4.3 after Lemma 4.5 and because the tilde model is semicontinuous, there is some $a^* \in \mathbf{A}^G$ satisfying

$$\begin{aligned} &\int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,a^*) - (q_x(a^*) - c^G(x,a^*))W^*(x) \\ = &\inf_{a\in\mathbf{A}^G}\left\{\int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))W^*(x)\right\} = 0, \end{aligned}$$

and hence $\int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,a^*) = (q_x(a^*) - c^G(x,a^*))W^*(x)$. This implies $q_x(a^*) \geq c^G(x,a^*)$ as the left hand side of the previous equality is nonnegative and $W^*(x) \geq 1$, and for the same reason, if $c^G(x,a^*) = q_x(a^*)$, then $c^G(x,a^*) = q_x(a^*) = 0$, in

which case,

$$
\begin{aligned}
& W^*(x) \geq 1 \\
= & \int_0^\infty \int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,a^*)e^{-\int_0^t (q_x(a^*)-c^G(x,a^*))ds}dt + e^{-\int_0^\infty q_x(a^*)ds}e^{\int_0^\infty c^G(x,a^*)ds} \\
\geq & \inf_{\hat{a}\in\hat{\mathbf{A}}} \left\{ \int_0^c \int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds}dt \right. \\
& +I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds} \\
& \left. +I\{c<\infty\}e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds} \int_{\mathbf{S}} W^*(y)e^{c^I(x,b,y)}Q(dy|x,b) \right\}.
\end{aligned}
$$

That is, (4.17) is satisfied by W^* at x, as desired. Finally, if $c^G(x,a^*) < q_x(a^*)$, then

$$
\begin{aligned}
& \inf_{\hat{a}\in\hat{\mathbf{A}}} \left\{ \int_0^c \int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds}dt \right. \\
& +I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds} \\
& \left. +I\{c<\infty\}e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds} \int_{\mathbf{S}} W^*(y)e^{c^I(x,b,y)}Q(dy|x,b) \right\} \\
\leq & \int_0^\infty \int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,a^*)e^{-\int_0^t (q_x(a^*)-c^G(x,a^*))ds}dt + e^{-\int_0^\infty q_x(a^*)ds}e^{\int_0^\infty c^G(x,a^*)ds} \\
= & \frac{\int_{\mathbf{S}} W^*(y)\tilde{q}(dy|x,a^*)}{q_x(a^*)-c^G(x,a^*)} + 0 = W^*(x),
\end{aligned}
$$

as requested. Thus, W^* satisfies (4.17). Consequently, $W^* = V^*$ on $\mathbf{S}$, as required. □

Proof of Theorem 4.1. Part (b) was seen in the proof of Lemma 4.4.

Consider the pair of measurable mappings (ψ^*, f^*) from Proposition 4.1. Recall that $W^* = V^*$ on $\mathbf{S}$ by Lemma 4.6. Keeping in mind Remark 4.3, an inspection of the proof of Lemma 4.6 reveals that the deterministic stationary strategy $(\varphi(x), \psi^*(x), t \to \delta_{f^*(x)}(da)) \in \hat{\mathbf{A}}$ in the hat DTMDP model, where φ is

defined in part (c) of this theorem, attains the infimum in

$$\begin{aligned} V^*(x) &= \inf_{\hat{a}\in\hat{\mathbf{A}}} \Big\{ \int_0^c \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,\rho_t) e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds} dt \\ &\quad + I\{c=\infty\} e^{-\int_0^\infty q_x(\rho_s)ds} e^{\int_0^\infty c^G(x,\rho_s)ds} \\ &\quad + I\{c<\infty\} e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds} \int_{\mathbf{S}} V^*(y) e^{c^I(x,b,y)} Q(dy|x,b) \Big\} \end{aligned}$$

for each $x \in \mathbf{S}$. By Lemma 4.1, this deterministic stationary strategy $(\varphi(x), \psi^*(x), t \to \delta_{f^*(x)}(da)) \in \hat{\mathbf{A}}$ is optimal for problem (4.12) for the hat DTMDP model. This and Remark 4.2 imply that $V^* = \mathcal{L}^*$ on $\mathbf{S}$ and part (c). By Lemma 4.6, we see now $\mathcal{L}^* = W^*$ on $\mathbf{S}$, and thus part (a) holds. □

Proof of Corollary 4.1. This corollary follows at once from Theorem 4.1, Lemma 4.5 and Remark 4.3. □

5 Risk-sensitive PDMDP with nonnegative cost rates

In this chapter we consider a piecewise deterministic Markov decision process (PDMDP), where the expected exponential utility of total (nonnegative) cost is to be minimized. The cost rate, transition rate and post-jump distributions are under control. The state space is Borel, and the transition and cost rates are locally integrable along the drift. Under natural conditions, we establish the optimality equation, justify the value iteration algorithm, and show the existence of a deterministic stationary optimal policy. Applied to special cases, the obtained results already significantly improve some existing results in the literature on finite horizon and infinite horizon discounted risk-sensitive continuous-time Markov decision processes.

Between two consecutive jumps, the state of the process evolves according to a measurable mapping ϕ from $\mathbf{S}\times[0,\infty)$ to $\mathbf{S}$, see (5.2) below. It is assumed that for each $x\in\mathbf{S}$

$$\phi(x,t+s)=\phi(\phi(x,t),s),\ \forall\ s,t\geq 0;\ \phi(x,0)=x, \tag{5.1}$$

and $t\to\phi(x,t)$ is continuous.

The marked point process $\{t_n,x_n\}$ defines the stochastic process $\{\xi_t,t\geq 0\}$ on $(\Omega,\mathcal{F})$ of interest by

$$\xi_t=\sum_{n\geq 0}I\{t_n\leq t<t_{n+1}\}\phi(x_n,t-t_n)+I\{t_\infty\leq t\}x_\infty,\ t\geq 0, \tag{5.2}$$

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

Definition 5.1. (The risk-sensitive PDMDP criterion)

For each $x \in \mathbf{S}$, and policy $\pi = (\pi_n)$,

$$V(x,\pi) := E_x^\pi \left[e^{\int_0^\infty \int_A c(\xi_t,a)\pi(da|\omega,t)dt} \right] = E_x^\pi \left[e^{\sum_{n=0}^\infty \int_0^{\theta_{n+1}} \int_{\mathbf{A}} c(\phi(x_n,s),a)\pi_n(da|x_0,\theta_1,\ldots,x_n,s)ds} \right]$$

A policy π^* is called optimal if for each $x \in \mathbf{S}$

$$V(x,\pi^*) = \inf_{\pi\in\Pi} V(x,\pi) =: V^*(x) \tag{5.3}$$

Here and below, we put $c(x_\infty, a) := 0$ for each $a \in \mathbf{A}$, and $\phi(x_\infty, t) = x_\infty$ for each $t \in [0,\infty)$.

Finally let the cost rate c be a $[0,\infty)$-valued measurable function on $\mathbf{S} \times \mathbf{A}$. For simplicity, we do not consider the case of different admissible action spaces at different states.

Condition 5.1. (a) For each bounded measurable function f on $\mathbf{S}$ and each $x \in \mathbf{S}$, $\int_{\mathbf{S}} f(y)\tilde{q}(dy|x,a)$ is continuous in $a \in \mathbf{A}$.

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

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

Condition 5.2. For each $x \in \mathbf{S}$, $\int_0^t \overline{q}_{\phi(x,s)}ds < \infty$, and $\int_0^t \sup_{a\in\mathbf{A}} c(\phi(x,s),a)ds < \infty$, for each $t \in [0,\infty)$.

The integrals in the above condition are well defined: the integrands are universally measurable in $s \in [0,\infty)$; see Chapter 7 of [9].

Roughly speaking, the uncontrolled version of the process evolves as follows: given the current state, the process evolves deterministically according to the mapping ϕ, up to the next jump, taking place after a random time whose distribution is (nonstationary) exponential, and the dynamics continue in the similar manner. A detailed book treatment with many examples of this and more general type of processes, allowing deterministic jumps, can be found in [23].

The objective of this chapter is to show, under the imposed conditions, the existence of a deterministic stationary optimal policy, and to establish the corresponding optimality equation satisfied by the value function V^*, together with its value iteration. Evidently, $V^*(x) \geq 1$ for each $x \in S$. Under the next condition, it will be seen that for each $x \in \mathbf{S}$, $V^*(\phi(x,s))$ is absolutely continuous in s.

Condition 5.3. For each $x \in \mathbf{S}$, $V^*(x) < \infty$.

The above condition is mainly assumed for notational convenience. In fact, the main optimality results (such as the existence of a deterministic stationary optimal policy) obtained in this paper can be established without assuming Condition 5.3, at the cost of some additional notations. In a nutshell, one has to consider the sets $\hat{\mathbf{S}} := \{x \in \mathbf{S} : \ V^*(x) < \infty\}$ and $\mathbf{S} \setminus \hat{\mathbf{S}}$ separately, and note that if $x \in \hat{\mathbf{S}}$, then $\phi(x,t) \in \hat{\mathbf{S}}$ for each $t \in [0,\infty)$. The reasoning presented under Condition 5.3 can be followed in an obvious manner. We formulate the corresponding optimality results in Remarks 5.1 and 5.2 below.

5.1 Main statements

We first present the main optimality results concerning problem (5.3) for the PDMDP model. Their proofs are postponed to the next section. Here and below, we assume that $q_{\phi(x,t)}(a) > \varepsilon(x) > 0$. This additional assumption is because that we can find the examples in Chapter 4 saying that when $q_x(a) = 0$, one of the continuity condition $\int_{\boldsymbol{X}} f(z)p(dz|(\theta,x),a)$ *is continuous for each bounded measurable function* f *on* $\boldsymbol{X}$ does not always hold, see Example 4.3 and Example 4.4.

Theorem 5.1. Suppose Conditions 5.1, 5.2 and 5.3 are satisfied. Then the following assertions hold.

(a) The value function V^* for problem (5.3) is the minimal $[1,\infty)$-valued solu-

tion to the following optimality equation:

$$\begin{aligned} & -(V(\phi(x,t)) - V(x)) \\ = & \int_0^t \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} V(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V(\phi(x,\tau)) \right\} d\tau, \\ & t \in [0,\infty), x \in \mathbf{S}. \end{aligned}$$

In particular, $V^*(\phi(x,t))$ is absolutely continuous in t for each $x \in \mathbf{S}$.

(b) There exists a deterministic stationary optimal policy f, which can be taken as any measurable mapping from $\mathbf{S}$ to $\mathbf{A}$ such that

$$\begin{aligned} & \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x)) \right\} \\ = & \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,f(x)) - (q_x(f(x)) - c(x,f(x)))V^*(x)), \ \forall\ x \in \mathbf{S}. \end{aligned}$$

Remark 5.1. By inspecting its proof, one can see the following version of Theorem 5.1 holds without assuming Condition 5.3. Suppose Conditions 5.1 and 5.2 are satisfied. Then the following assertions hold.

(a) The value function V^* for problem (5.3) is the minimal $[1,\infty]$-valued solution to the following optimality equation:

$$\begin{aligned} & -(V(\phi(x,t)) - V(x)) \\ = & \int_0^t \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} V(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V(\phi(x,\tau)) \right\} d\tau, \\ & t \in [0,\infty), x \in \hat{\mathbf{S}}; \\ & V(x) < \infty,\ x \in \hat{\mathbf{S}};\ V(x) = \infty,\ x \in \mathbf{S} \setminus \hat{\mathbf{S}}. \end{aligned}$$

In particular, $V^*(\phi(x,t))$ is absolutely continuous in t for each $x \in \hat{\mathbf{S}}$.

(b) There exists a deterministic stationary optimal policy f, which can be taken

as any measurable mapping from $\mathbf{S}$ to $\mathbf{A}$ such that

$$\begin{aligned} & \inf_{a\in\mathbf{A}}\left\{\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x))\right\} \\ = & \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,f(x)) - (q_x(f(x)) - c(x,f(x)))V^*(x)),\ \forall\ x\in\hat{\mathbf{S}}. \end{aligned}$$

Next, we present the value iteration algorithm for the value function V^*.

Theorem 5.2. Suppose Conditions 5.1, 5.2 and 5.3 are satisfied. Let $V^{(0)}(x) = 1$ for each $x \in \mathbf{S}$. For each $n \geq 0$, let $V^{(n+1)}$ be the minimal $[1,\infty)$-valued measurable solution to

$$\begin{aligned} & -(V^{(n+1)}(\phi(x,t)) - V^{(n+1)}(x)) \\ = & \int_0^t \inf_{a\in\mathbf{A}}\left\{\int_{\mathbf{S}} V^{(n)}(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V^{(n+1)}(\phi(x,\tau))\right\} d\tau, \\ & t\in[0,\infty), x\in\mathbf{S}, \end{aligned} \tag{5.4}$$

such that $V^{(n+1)}(\phi(x,t))$ is absolutely continuous in t for each $x \in \mathbf{S}$. (For each $n \geq 0$, such a solution always exists.) Furthermore, $\{V^{(n)}\}$ is a monotone nondecreasing sequence of measurable functions on $\mathbf{S}$ such that for each $x \in \mathbf{S}$, $V^{(n)}(x) \uparrow V^*(x)$ as $n \uparrow \infty$.

Remark 5.2. Similar to Remark 5.1, we have the following version of Theorem 5.2 without assuming Condition 5.3. Suppose Conditions 5.1, 5.2 are satisfied. Let $V^{(0)}(x) = 1$ for each $x \in \hat{\mathbf{S}}$ and $V^{(0)}(x) = \infty$ if $x \in \mathbf{S}\setminus\hat{\mathbf{S}}$. For each $n \geq 0$, let $V^{(n+1)}$ be the minimal $[1,\infty]$-valued measurable solution to

$$\begin{aligned} & -(V^{(n+1)}(\phi(x,t)) - V^{(n+1)}(x)) \\ = & \int_0^t \inf_{a\in\mathbf{A}}\left\{\int_{\mathbf{S}} V^{(n)}(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V^{(n+1)}(\phi(x,\tau))\right\} d\tau, \\ & t\in[0,\infty), x\in\hat{\mathbf{S}}, \\ & V^{(n+1)}(x) < \infty,\ x\in\hat{\mathbf{S}},\ V^{(n+1)}(x) = \infty,\ x\in\mathbf{S}\setminus\hat{\mathbf{S}}. \end{aligned}$$

Here $V^{(n+1)}(\phi(x,t))$ is absolutely continuous in t for each $x \in \hat{\mathbf{S}}$. (For each $n \geq 0$, such a solution always exists.) Furthermore, $\{V^{(n)}\}$ is a monotone nondecreasing sequence of measurable functions on $\mathbf{S}$ such that for each $x \in \mathbf{S}$, $V^{(n)}(x) \uparrow V^*(x)$

as $n \uparrow \infty$.

We can apply our theorems to a special case of a CTMDP. That is, $\phi(x,t) \equiv x$ for each $x \in \mathbf{S}$. We next give two applications of what we obtained for PDMDP model, which mainly focus on the transformation between them.

The first application considering the following α-discounted risk-sensitive CTMDP problem was considered in [43]:

$$\text{Minimize over } \pi \in \Pi\text{: } E_x^\pi \left[e^{\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c(\xi_t,a)\pi(da|\omega,t)dt} \right],\ x \in \mathbf{S}. \tag{5.5}$$

Here $\alpha > 0$ is a fixed constant. In fact, the authors of [43] were restricted to Markov policies, bounded transition and cost rates, i.e., $\sup_{x\in\mathbf{S}} \overline{q}_x < \infty$, and $\sup_{x\in\mathbf{S},a\in\mathbf{A}} c(x,a) < \infty$, and a finite state space $\mathbf{S}$. These restrictions were needed for their investigations, see e.g., Remark 3.6 in [43]. Under the compactness-continuity condition (Condition 5.1), it was shown in [43] that there exists an optimal Markov policy for the discounted risk-sensitive CTMDP, and established the optimality equation. By using the theorems presented earlier in this section, we can obtain these optimality results for problem (5.5) in a much more general setup: the state space $\mathbf{S}$ is Borel, there is no boundedness requirement on the transition rate with respect to the state $x \in \mathbf{S}$, and the optimality is over the class of history-dependent policies. Furthermore, we let the CTMDP model be nonhomogeneous, i.e., the transition rate $q(dy|t,x,a)$ now is a signed kernel on $\mathcal{B}(\mathbf{S})$ from $(t,x,a) \in [0,\infty)\times\mathbf{S}\times\mathbf{A}$, satisfying the corresponding version of (1.4); the notations $\tilde{q}$ is kept as before, with the extra argument t in addition to x. Similarly, the nonnegative cost rate c is allowed to be a measurable function on $[0,\infty)\times\mathbf{S}\times\mathbf{A}$.

Corollary 5.1. Consider the α-discounted risk-sensitive (nonhomogeneous) CTMDP problem (5.5) with $c(\xi_t,a)$ being replaced by $c(t,\xi_t,a)$. Suppose

$$\sup_{t\in[0,\infty)} \{\overline{q}_{(t,x)}\} < \infty,\ \forall\ x \in \mathbf{S}, \qquad \sup_{t\in[0,\infty),x\in\mathbf{S},a\in\mathbf{A}} c(t,x,a) < \infty,$$

and the corresponding version of Condition 5.1, where x is replaced by (t, x), is satisfied by the nonhomogeneous CTMDP model. Then the following assertions hold.

(a) There exists some $[1, \infty)$-valued measurable solution on $[0, \infty) \times \mathbf{S}$ to

$$\begin{aligned} & -(V(t,x) - V(0,x)) \\ = & \int_0^t \inf_{a \in \mathbf{A}} \left\{ \int_{\mathbf{S}} V(u,y) \tilde{q}(dy|u,x,a) + (e^{-\alpha u} c(u,x,a) - q_{(u,x)}(a)) V(u,x) \right\} du, \\ & x \in \mathbf{S},\ t \in [0,\infty), \end{aligned}$$

where $V(t, x)$ and $V(0, x)$ correspond to the $V(\phi(t, x))$ and $V(x)$ in PDMDP respectively and V is actually the criterion of this application model (5.5). $V(t, x)$ is absolutely continuous in t for each $x \in \mathbf{S}$.

(b) Let L be the minimal $[1, \infty)$-valued measurable solution on $[0, \infty) \times \mathbf{S}$ to the above equation. Then the value function say L^* to the α-discounted risk-sensitive CTMDP problem (5.5) (with $c(\xi_t, a)$ being replaced by $c(t, \xi_t, a)$) is given by $L^*(x) = L(0, x)$ for each $x \in \mathbf{S}$.

(c) There exists an optimal deterministsic Markov policy f for the α-discounted risk-sensitive CTMDP problem (5.5) (with $c(\xi_t, a)$ being replaced by $c(t, \xi_t, a)$). One can take f as any measurable mapping from $[0, \infty) \times \mathbf{S}$ to $\mathbf{A}$ such that

$$\begin{aligned} & \inf_{a \in \mathbf{A}} \left\{ \int_{\mathbf{S}} L(u,y) \tilde{q}(dy|u,x,a) + (e^{-\alpha u} c(u,x,a) - q_{(u,x)}(a)) L(u,x) \right\} \\ = & \int_{\mathbf{S}} L(u,y) \tilde{q}(dy|u,x,f(u,x)) + (e^{-\alpha u} c(u,x,f(u,x)) - q_{(u,x)}(f(u,x))) L(u,x) \end{aligned}$$

for each $u \in [0, \infty)$ and $x \in \mathbf{S}$.

Proof. We prove this by reformulating the nonhomogeneous version of the α-discounted risk-sensitive (nonhomogeneous) CTMDP problem (5.5) in the form of problem (5.3) for a PDMDP, which we introduce as follows. We use the notation

"hat" to distinguish this model from the original (nonhomogeneous) CTMDP model.

- The state space is $\hat{\mathbf{S}} = [0, \infty) \times \mathbf{S}$.
- The action space is the same as in the CTMDP: $\hat{\mathbf{A}} = \mathbf{A}$.
- the transition rate $\hat{q}(ds \times dy|(t,x),a)$ is defined by

  $$\hat{q}(ds \times dy|(t,x),a) := \tilde{\hat{q}}(ds \times dy|(t,x),a) - I\{(t,x) \in ds \times dy\} q_{(t,x)}(a),$$

  where

  $$\tilde{\hat{q}}(ds \times dy|(t,x),a) := I\{t \in ds\} \tilde{q}(dy|t,x,a),$$

  for each $(t,x) \in \hat{\mathbf{S}}$ and $a \in \hat{\mathbf{A}}$.
- The drift is given by $\hat{\phi}((t,x),s) := (t+s,x)$ for each $x \in \mathbf{S}$ and $t, s \geq 0$. Clearly it satisfies the corresponding version of (5.1).
- The cost rate is given by

  $$\hat{c}((t,x),a) := e^{-\alpha t} c(t,x,a), \ \forall\, t \in [0,\infty), \ x \in \mathbf{S}, \ a \in \mathbf{A}.$$

Now the marked point process $\{\hat{t}_n, \hat{x}_n\}$ and controlled process $\hat{\xi}_t$ in this PDMDP model is connected to those in the original (nonhomogeneous) CTMDP model, namely (t_n, x_n) and ξ_t, via $\hat{t}_n = t_n$ and $\hat{x}_n = (t_n, x_n)$, and $\hat{\xi}_t = (t, \xi_t)$.

Clearly, Conditions 5.1, 5.2 and 5.3 are satisfied by this PDMDP model. It remains to apply Theorem 5.1. □

The condition in the previous corollary is much weaker than in [43], and can be further weakened; one only needs the reformulated PDMDP to satisfy Conditions 5.1, 5.2 and 5.3. Moreover, the finiteness of the cost rate c was assumed in the previous corollary only to ensure Condition 5.3 to be satisfied. It can be relaxed

if one formulates the previous corollary using the statements in Remarks 5.1 and 5.2.

Here comes the second application that one can consider the α-discounted risk-sensitive nonhomogeneous CTMDP problem on the finite horizon $[0, T]$ with $T > 0$ being a fixed constant:

$$\text{Minimize over } \pi \in \Pi\text{: } E_x^\pi \left[e^{\int_0^T e^{-\alpha t} \int_{\mathbf{A}} c(t,\xi_t,a)\pi(da|\omega,t)dt + g(\xi_T)} \right], \ x \in \mathbf{S},$$

where g is a $[0, \infty)$-valued measurable function; $g(x)$ represents the terminal cost incurred when $\xi_T = x \in \mathbf{S}$. Let us put $g(x_\infty) := 0$. Here α is a fixed nonnegative finite constant. A simpler version of this problem was considered in [101] with $\alpha = 0$ and a bounded cost rate, where additional restrictions were put on the growth of the transition rate. We can reformulate this problem into the PDMDP problem (5.3) just as in the above but we add one more parameter $\Delta \in \{0, 1\}$. Now the 'prime' model is as below.

- The state space is $\mathbf{S}' = [0, \infty) \times \mathbf{S} \times \{0, 1\}$.
- The action space is the same: $\mathbf{A}' = \mathbf{A}$.
- the transition rate $q'(ds \times dy \times d\Delta|(t, x, \Delta), a)$ is defined by

$$\begin{aligned}
&q'(ds \times dy \times d\Delta|(t, x, 0), a) := \delta_0(d\Delta)\hat{q}(ds \times dy|(t, x), a) \text{ if t} \leq \text{T} \\
&q'(ds \times dy \times d\Delta|(t, x, 0), a) := \delta_1(d\Delta)\delta_{(t,x)}(ds \times dy) \text{ if t} > \text{T} \\
&q'(ds \times dy \times d\Delta|(t, x, 1), a) := \delta_0(d\Delta)\delta_{(t,x)}(ds \times dy)
\end{aligned}$$

- The drift is given by $\phi'((t, x, \Delta), s) := (t + s, x)$ for each $x \in \mathbf{S}$
- The cost rate is given by

$$c'((t, x, \Delta), a) = \begin{cases} e^{-\alpha t} c(t, x, a), & \text{if } t \leq T; \\ e^{-(t-T)} g(x) & \text{if } t > T. \end{cases}$$

Actually, we can notice that when $t \leq T$, it is the same as the above ($\Delta = 0$ in this case). But when $t \geq T$, we have to construct a "dynamic absorbing" system between $(t, x, 0)$ and $(t, x, 1)$ to make sure whenever the state attains one of them, it will jump to the other state with probability one. It is easy to see this $\mathcal{M}'$ model also satisfies Conditions 5.1 5.2.

5.2 Proof of the main statements

Lemma 5.1. Suppose Conditions 5.1 and 5.2 are satisfied. Then the following assertions hold.

(a) The value function V^* is the minimal $[1, \infty]$-valued measurable solution to

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

(b) The mapping

$$\begin{aligned} \rho \in \mathcal{R} \ \rightarrow \ W(x,\rho) &:= \int_0^\infty e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \left(\int_{\mathbf{S}} V^*(y) \tilde{q}(dy|\phi(x,\tau),\rho_\tau) \right) d\tau \\ &\quad + e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds} e^{\int_0^\infty c(\phi(x,s),\rho_s)ds} \end{aligned}$$

is lower semicontinuous for each $x \in \mathbf{S}$.

Proof. One can legitimately consider the following DTMDP (discrete-time Markov decision process): according to Lemma 2.29 of [19], all the involved mappings are measurable.

- The state space is $\mathbf{X} := ((0, \infty) \times \mathbf{S}) \bigcup \{(\infty, x_\infty)\}$. Whenever the topology is concerned, (∞, x_∞) is regarded as an isolated point in $\mathbf{X}$.
- The action space is $\mathbb{A} := \mathcal{R}$.

- The transition kernel p on $\mathcal{B}(\mathbf{X})$ from $\mathbf{X} \times \mathbb{A}$, is given for each $\rho \in \mathbb{A}$ by

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

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

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

The relevant facts and statements for the DTMDP are included in the Appendix.

One can show that under Conditions 5.1 and 5.2, for each $(\theta, x) \in \mathbf{X}$, $a \in \mathbb{A} \to \int_{\mathbf{X}} f(z)p(dz|(\theta, x), a)$ is continuous for each bounded measurable function f on $\mathbf{X}$; for each $(\theta, x) \in \mathbf{X}$ and $(\tau, y) \in \mathbf{X}$, $a \in \mathbb{A} \mapsto l((\theta, x), \rho, (\tau, y))$ is lower semicontinuous, and $\mathbb{A}$ is a compact Borel space. Hence, Condition B.2 for the DTMDP model $\{\mathbf{X}, \mathbb{A}, p, l\}$ is satisfied.

The controlled process in the above DTMDP model $\{\mathbf{X}, \mathbb{A}, p, l\}$ 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 PDMDP, Θ_0 is fictitious, and X_0 is the initial state in the PDMDP. Let Σ be the class of all strategies for the DTMDP model $\{\mathbf{X}, \mathbb{A}, p, l\}$, and Σ^0_{DM} be the class of deterministic Markov strategies in the form $\sigma = (\varphi_n)$ where $\varphi_0((\theta, x))$ does not depend on $\theta \in (0,\infty)$ for each $x \in \mathbf{S}$. We preserve the term of policy for the PDMDP and the term of strategy for the DTMDP.

According to Proposition B.1, the function

$$(\theta, x) \in \mathbf{X} \to \mathbf{V}^*((\theta, x)) := \inf_{\sigma \in \Sigma} \mathbb{E}^{\sigma}_{(\theta,x)} \left[e^{\sum_{n=0}^{\infty} l(Y_n, A_n, Y_{n+1})} \right]$$

is the minimal $[1, \infty]$-valued measurable solution to the optimality equation

$$\begin{aligned} \mathbf{V}^*((\theta, x)) \;\; = \;\; & \inf_{\rho \in \mathcal{R}} \left\{ \int_0^{\infty} e^{-\int_0^{\tau} (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \left(\int_{\mathbf{S}} \mathbf{V}^*((\tau, y)) \tilde{q}(dy | \phi(x,\tau), \rho_\tau) \right) d\tau \right. \\ & \left. + e^{-\int_0^{\infty} q_{\phi(x,s)}(\rho_s) ds} e^{\int_0^{\infty} c(\phi(x,s),\rho_s) ds} \right\} \end{aligned}$$

for each $x \in \mathbf{S}$ and $\theta \in (0, \infty)$; this is just (B.3). Furthermore, by Proposition B.1, there exists a deterministic stationary strategy σ^* for the DTMDP such that $\sigma^*((\theta, x))$ attains the above infimum for each $x \in \mathbf{S}$ and $\theta \in (0, \infty)$, and any such strategy σ^* verifies

$$\mathbb{E}^{\sigma^*}_{(\theta,x)} \left[e^{\sum_{n=0}^{\infty} l(Y_n, A_n, Y_{n+1})} \right] = \inf_{\sigma \in \Sigma} \mathbb{E}^{\sigma}_{(\theta,x)} \left[e^{\sum_{n=0}^{\infty} l(Y_n, A_n, Y_{n+1})} \right], \; \forall \; (\theta, x) \in \mathbf{X}.$$

Let $\hat{\theta} \in \; (0, \infty)$ be arbitrarily fixed. The function $\mathbf{V}^*((\theta, x))$ being measurable in $(\theta, x) \in \mathbf{X}$, it follows that $x \in \mathbf{S} \to \mathbf{V}^*((\hat{\theta}, x))$ is measurable. The strategy σ^* and the constant $\hat{\theta}$ induce a deterministic Markov strategy $\sigma^{**} = (\varphi_n) \in \Sigma^0_{DM}$, where $\varphi_0((\theta, x)) =: \sigma^*((\hat{\theta}, x))$ for each $\theta \in (0, \infty), \; x \in \mathbf{S}$, and $\varphi_n((\theta, x)) := \sigma^*((\theta, x))$ for each $n \geq 1, \; \theta \in (0, \infty), \; x \in \mathbf{S}$. (The control on the isolated point $(0, x_\infty)$ is irrelevant and we do not specify the definition of the strategy on that point.) This strategy can be identified with a policy π^* in the PDMDP. On the other hand, each policy $\pi = (\pi_n)$ can be identified with a deterministic strategy in this DTMDP. Thus,

$$\begin{aligned} & V^*(x) \geq \mathbf{V}^*((\hat{\theta}, x)) = \mathbb{E}^{\sigma^*}_{(\hat{\theta},x)} \left[e^{\sum_{n=0}^{\infty} l(Y_n, A_n, Y_{n+1})} \right] \\ = \;\; & \mathbb{E}^{\sigma^{**}}_{(\hat{\theta},x)} \left[e^{\sum_{n=0}^{\infty} l(Y_n, A_n, Y_{n+1})} \right] = V(x, \pi^*) \geq V^*(x) \end{aligned}$$

for each $x \in \mathbf{S}$. Consequently, the policy π^* is optimal, $V^*(x) = \mathbf{V}^*((\hat{\theta}, x))$ for each $x \in \mathbf{S}$ and $\hat{\theta} \in (0, \infty)$; recall that $\hat{\theta}$ was arbitrarily fixed. The statement of

this lemma now follows. □

The policy π^* in the proof of the previous lemma is actually optimal for problem (5.3). However, it is not necessarily a deterministic nor stationary policy. Also the reduction of the risk-sensitive PDMDP problem (5.3) to a risk-sensitive problem for the DTMDP model $\{\mathbf{X}, \mathbb{A}, p, l\}$ as seen in the proof of the above theorem will be used without special reference in what follows.

Lemma 5.2. Suppose Conditions 5.1, 5.2 and 5.3 are satisfied. For each $x \in \mathbf{S}$ and $\rho \in \mathcal{R}$,

$$\begin{aligned} t \in [0,\infty) \quad \to \quad & \int_0^t e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)d\tau \\ & + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} V^*(\phi(x,t)) \end{aligned}$$

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

Proof. Let $0 \le t_1 < t_2 < \infty$ be arbitrarily fixed. We need show

$$\begin{aligned} & \int_0^{t_2} e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)d\tau \\ & + e^{-\int_0^{t_2} (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} V^*(\phi(x,t_2)) \\ \ge \quad & \int_0^{t_1} e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)d\tau \\ & + e^{-\int_0^{t_1} (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} V^*(\phi(x,t_1)). \end{aligned} \tag{5.6}$$

It is without loss of generality to assume

$$\int_0^{t_2} e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)d\tau < \infty.$$

Then all the four terms in (5.6) are nonnegative and finite, and (5.6) is equivalent

to

$$
\begin{aligned}
& \int_0^{t_2} e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)d\tau \\
& +e^{-\int_0^{t_2}(q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds}V^*(\phi(x,t_2)) \\
& -\int_0^{t_1} e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)d\tau \\
& -e^{-\int_0^{t_1}(q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds}V^*(\phi(x,t_1)) \\
= & \int_{t_1}^{t_2} e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)d\tau \\
& +e^{-\int_0^{t_1}(q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds}\left(e^{-\int_{t_1}^{t_2}(q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds}V^*(\phi(x,t_2))-V^*(\phi(x,t_1))\right) \\
= & \left\{\int_0^{t_2-t_1} e^{-\int_0^\tau (q_{\phi(x,s+t_1)}(\rho_{s+t_1})-c(\phi(x,s+t_1),\rho_{s+t_1}))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,t_1+\tau),\rho_{t_1+\tau})d\tau\right. \\
& \left.+e^{-\int_{t_1}^{t_2}(q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds}V^*(\phi(x,t_2))-V^*(\phi(x,t_1))\right\} e^{-\int_0^{t_1}(q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds} \\
\geq & 0, \qquad (5.7)
\end{aligned}
$$

which is verified as follows. Let $\delta > 0$ be arbitrarily fixed. By Lemma 5.1, there exists some $\hat{\nu} \in \mathcal{R}$ such that

$$
\begin{aligned}
V^*(\phi(x,t_2))+\delta \geq & \int_0^\infty \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,t_2+\tau),\hat{\nu}_\tau)e^{-\int_0^\tau (q_{\phi(x,t_2+s)}(\hat{\nu}_s)-c(\phi(x,t_2+s),\hat{\nu}_s))ds}d\tau \\
& +e^{-\int_0^\infty q_{\phi(x,t_2+s)}(\hat{\nu}_s)ds}e^{\int_0^\infty c(\phi(x,t_2+s),\hat{\nu}_s)ds}.
\end{aligned}
$$

(Recall $\phi(x,t_2+t)=\phi(\phi(x,t_2),t)$ for each $t \geq 0$.) Consider $\tilde{\nu} \in \mathcal{R}$ defined by

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

Then routine calculations lead to

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

Since $\delta > 0$ was arbitrarily fixed, now it follows that the term in the parenthesis in (5.7) is nonnegative, and thus inequality (5.7) is verified. □

Lemma 5.3. Suppose Conditions 5.1, 5.2 and 5.3 are satisfied. For each $x \in \mathbf{S}$, there is some $\rho^* \in \mathcal{R}$ such that

$$
\begin{aligned}
V^*(x) &= \inf_{\rho\in\mathcal{R}} \left\{ \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho_v)-c(\phi(x,v),\rho_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),\rho_s)ds \right. \\
&\quad \left. + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds} V^*(\phi(x,t)) \right\} \\
&= \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho_v^*)-c(\phi(x,v),\rho_v^*))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),\rho_s^*)ds \\
&\quad + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s^*)-c(\phi(x,s),\rho_s^*))ds} V^*(\phi(x,t)),\ \forall\ t \geq 0. \qquad (5.8)
\end{aligned}
$$

Proof. Let $x \in \mathbf{S}$ be fixed, and let $\rho^* \in \mathcal{R}$ be such that $V^*(x) = W(x,\rho^*)$, see Lemma 5.1. Suppose $t \in [0,\infty)$ is arbitrarily fixed. Consider $\tilde{\rho} \in \mathcal{R}$ defined by

$\tilde{\rho}_s = \rho^*_{t+s}$ for each $s > 0$. Then

$$\begin{aligned} V^*(x) &= \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho^*_v) - c(\phi(x,v),\rho^*_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),\rho^*_s)ds \\ &+ e^{-\int_0^t (q_{\phi(x,s)}(\rho^*_s) - c(\phi(x,s),\rho^*_s))ds} \times \left\{ \int_0^\infty e^{-\int_0^\tau (q_{\phi(x,t+s)}(\tilde{\rho}_s) - c(\phi(x,s+t),\tilde{\rho}_s))ds} \right. \\ &\left. \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau+t),\tilde{\rho}_\tau)d\tau + e^{-\int_0^\infty q_{\phi(x,t+s)}(\tilde{\rho}_s)ds} e^{-\int_0^\infty c(\phi(x,t+s),\tilde{\rho}_s)ds} \right\} \\ &\geq \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho^*_v) - c(\phi(x,v),\rho^*_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),\rho^*_s)ds \\ &+ e^{-\int_0^t (q_{\phi(x,s)}(\rho^*_s) - c(\phi(x,s),\rho^*_s))ds} V^*(\phi(x,t)); \end{aligned}$$

recall (5.1). On the other hand, by Lemma 5.2,

$$\begin{aligned} V^*(x) \leq \inf_{\rho\in\mathcal{R}} \left\{ \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho_v) - c(\phi(x,v),\rho_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),\rho_s)ds \right. \\ \left. + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} V^*(\phi(x,t)) \right\}. \end{aligned}$$

The statement of this lemma is thus proved. □

Lemma 5.4. Suppose Conditions 5.1, 5.2 and 5.3 are satisfied. Then for each $x \in \mathbf{S}$, $t \in [0,\infty) \to V^*(\phi(x,t))$ is absolutely continuous.

Proof. This immediately follows from Lemma 5.3 and (5.8). □

Proof of Theorem 5.1. (a) Under Conditions 5.1, 5.2 and 5.3, by Lemma 5.4, for each $x \in \mathbf{S}$, let $t \in [0,\infty) \to U^*(x,t)$ be an integrable real-valued function such that $U^*(x,t)$ coincides with the derivative of $t \in [0,\infty) \to V(\phi(x,t))$ almost everywhere, that is, $U^*(x,t) \triangleq \frac{dV(\phi(x,t))}{dt}$. Let $x \in \mathbf{S}$ and $t \in [0,\infty)$ be fixed, and let $\rho^* \in \mathcal{R}$ be from Lemma 5.3.

By Lemmas 5.3 and 5.4,

$$\int_0^\tau e^{-\int_0^s (q_{\phi(x,v)}(\rho^*_v) - c(\phi(x,v),\rho^*_v))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),\rho^*_s)ds$$

and

$$e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s^*)-c(\phi(x,s),\rho_s^*))ds}V^*(\phi(x,\tau))$$

are absolutely continuous in τ and are finite for each $\tau \in [0,\infty)$. Since $\phi(x,0)=x$, see (5.1),

$$\begin{aligned} & e^{-\int_0^t (q_{\phi(x,s)}(\rho_s^*)-c(\phi(x,s),\rho_s^*))ds}V^*(\phi(x,t)) - V^*(x) \\ = & \int_0^t e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s^*)-c(\phi(x,s),\rho_s^*))ds}\left\{U^*(x,\tau) - (q_{\phi(x,\tau)}(\rho_\tau^*) - c(\phi(x,\tau),\rho_\tau^*))V^*(\phi(x,\tau))\right\}d\tau. \end{aligned}$$

which, together with Lemma 5.3, gives

$$\begin{aligned} 0 = & \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho_v^*)-c(\phi(x,v),\rho_v^*))dv}\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),\rho_s^*)ds \\ & + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s^*)-c(\phi(x,s),\rho_s^*))ds}V^*(\phi(x,t)) - V^*(x) \\ = & \int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*)-c(\phi(x,v),\rho_v^*))dv}\left\{\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau^*) + U^*(x,\tau)\right. \\ & \left. -(q_{\phi(x,\tau)}(\rho_\tau^*) - c(\phi(x,\tau),\rho_\tau^*))V^*(\phi(x,\tau))\right\}d\tau \\ \geq & \int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*)-c(\phi(x,v),\rho_v^*))dv}\left\{U^*(x,\tau)\right. \\ & \left. + \inf_{a\in\mathbf{A}}\left\{\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V^*(\phi(x,\tau))\right\}\right\}d\tau \\ = & \int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*)-c(\phi(x,v),\rho_v^*))dv}\left\{U^*(x,\tau) + \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),f(\phi(x,\tau)))\right. \\ & \left. -(q_{\phi(x,\tau)}(f(\phi(x,\tau))) - c(\phi(x,\tau),f(\phi(x,\tau))))V^*(\phi(x,\tau))\right\}d\tau, \end{aligned} \tag{5.9}$$

where f is a measurable mapping from $\mathbf{S}$ to $\mathbf{A}$ such that

$$\begin{aligned} & \inf_{a\in\mathbf{A}}\left\{\int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V^*(x)\right\} \\ = & \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|x,f(x)) - (q_x(\varphi(x)) - c(x,f(x)))V^*(x) \end{aligned}$$

for each $x \in \mathbf{S}$; the existence of such a mapping is according to a well known measurable selection theorem, c.f. Proposition D.5 of [57].

Note that $e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v)-c(\phi(x,v),\rho_v))dv}$ is bounded and separated from zero in $\tau \in [0,t]$ for each $\rho \in \mathcal{R}$; recall Condition 5.2. So

$$\int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*)-c(\phi(x,v),\rho_v^*))dv} \left\{U^*(x,\tau) - (q_{\phi(x,\tau)}(f(\phi(x,\tau))) - c(\phi(x,\tau), f(\phi(x,\tau))))V^*(\phi(x,\tau))\right\} d\tau$$

is finite. If

$$\int_0^t \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau), f(\phi(x,\tau)))d\tau = \infty,$$

then

$$\begin{aligned}&\int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*)-c(\phi(x,v),\rho_v^*))dv} \left\{U^*(x,\tau) + \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau), f(\phi(x,\tau)))\right.\\ &\left.-(q_{\phi(x,\tau)}(f(\phi(x,\tau))) - c(\phi(x,\tau), f(\phi(x,\tau))))V^*(\phi(x,\tau))\right\} d\tau = \infty,\end{aligned}$$

which is against (5.9). Therefore,

$$\int_0^t \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau), f(\phi(x,\tau)))d\tau < \infty.$$

Then

$$\begin{aligned}&\int_0^v e^{-\int_0^\tau (q_{\phi(x,s)}(f(\phi(x,s)))-c(\phi(x,s),f(\phi(x,s))))ds} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau), f(\phi(x,\tau)))d\tau\\ &+e^{-\int_0^v (q_{\phi(x,s)}(f(\phi(x,s)))-c(\phi(x,s),f(\phi(x,s))))ds} V^*(\phi(x,v))\end{aligned}$$

is absolutely continuous on $[0,t]$. After legitimately differentiating the above expression with respect to v, and applying Lemma 5.2, we see

$$\begin{aligned}&U^*(x,v) + \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,v), f(\phi(x,v)))\\ &-(q_{\phi(x,v)}(f(\phi(x,v))) - c(\phi(x,v), f(\phi(x,v))))V^*(\phi(x,v)) \geq 0\end{aligned}$$

for almost all $v \in [0,t]$. This and (5.9) imply

$$U^*(x,\tau) + \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V^*(\phi(x,\tau)) \right\} = 0$$

almost everywhere in $\tau \in [0,t]$. Remember, $t \in [0,\infty)$ was arbitrarily fixed. The first part of (a) is thus verified, and we postpone the justification of the second part of (a) after the proof of part (b).

(b) We use the same notation as in the above. Note that

$$\liminf_{t\to\infty} \left\{ e^{-\int_0^t (q_{\phi(x,s)}(f(\phi(x,s))) - c(\phi(x,s),f(\phi(x,s))))ds} \right\} \geq e^{-\int_0^\infty q_{\phi(x,s)}(f(\phi(x,s)))ds} e^{\int_0^\infty c(\phi(x,s),f(\phi(x,s)))ds} \quad (5.10)$$

Indeed, if either $\int_0^\infty q_{\phi(x,s)}(f(\phi(x,s)))ds$ or $\int_0^\infty c(\phi(x,s),f(\phi(x,s))))ds$ is finite, then in the above inequality, the equality takes place; if both $\int_0^\infty q_{\phi(x,s)}(f(\phi(x,s)))ds$ and $\int_0^\infty c(\phi(x,s),f(\phi(x,s))))ds$ are infinite, then the right hand side of the inequality is zero.

In the proof of part (a), it was observed that

$$\int_0^t e^{-\int_0^s (q_{\phi(x,v)}(f(\phi(x,v))) - c(\phi(x,v),f(\phi(x,v))))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),f(\phi(x,s)))ds$$

and

$$e^{-\int_0^t (q_{\phi(x,s)}(f(\phi(x,s))) - c(\phi(x,s),f(\phi(x,s))))ds} V^*(\phi(x,t))$$

are absolutely continuous in t and are thus finite for each $t \in [0,\infty)$. As in the proof of part (a), similar calculations to those in (5.9) imply that for each $t \in [0,\infty)$,

$$\begin{aligned}
&\int_0^t e^{-\int_0^s (q_{\phi(x,v)}(f(\phi(x,v))) - c(\phi(x,v),f(\phi(x,v))))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s),f(\phi(x,s)))ds \\
&+e^{-\int_0^t (q_{\phi(x,s)}(f(\phi(x,s))) - c(\phi(x,s),f(\phi(x,s))))ds} V^*(\phi(x,t)) - V^*(x) \\
= &\int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(f(\phi(x,v))) - c(\phi(x,v),f(\phi(x,v))))dv} \left\{ U^*(x,\tau) + \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,\tau),f(\phi(x,\tau))) \right. \\
&\left. -(q_{\phi(x,\tau)}(f(\phi(x,\tau))) - c(\phi(x,\tau),f(\phi(x,\tau))))V^*(\phi(x,\tau)) \right\} d\tau = 0,
\end{aligned}$$

where the last equality is by what was established in part (a). Therefore, for

each $t \in [0, \infty)$,

$$\begin{aligned} & V^*(x) - \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(f(\phi(x,v))) - c(\phi(x,v), f(\phi(x,v))))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s), f(\phi(x,s)))ds \\ = \; & e^{-\int_0^t (q_{\phi(x,s)}(f(\phi(x,s))) - c(\phi(x,s), f(\phi(x,s))))ds} V^*(\phi(x,t)) \\ \geq \; & e^{-\int_0^t (q_{\phi(x,s)}(f(\phi(x,s))) - c(\phi(x,s), f(\phi(x,s))))ds}, \end{aligned}$$

where the inequality holds because $V^*(x) \geq 1$ for each $x \in \mathbf{S}$. Taking $\liminf_{t\to\infty}$ on the both sides of the previous equality yields:

$$\begin{aligned} & V^*(x) - \int_0^\infty e^{-\int_0^s (q_{\phi(x,v)}(f(\phi(x,v))) - c(\phi(x,v), f(\phi(x,v))))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s), f(\phi(x,s)))ds \\ \geq \; & e^{-\int_0^\infty q_{\phi(x,s)}(f(\phi(x,s)))ds} e^{\int_0^\infty c(\phi(x,s), f(\phi(x,s))))ds} \end{aligned}$$

with the inequality following from (5.10). Hence

$$\begin{aligned} V^*(x) \; \geq \; & \int_0^\infty e^{-\int_0^s (q_{\phi(x,v)}(f(\phi(x,v))) - c(\phi(x,v), f(\phi(x,v))))dv} \int_{\mathbf{S}} V^*(y)\tilde{q}(dy|\phi(x,s), f(\phi(x,s)))ds \\ & + e^{-\int_0^\infty q_{\phi(x,s)}(f(\phi(x,s)))ds} e^{\int_0^\infty c(\phi(x,s), f(\phi(x,s))))ds} = W(x, \tilde{f}^x) \geq V^*(x). \end{aligned}$$

Here it is clear that $s \in [0, \infty) \to f(\phi(x,s))$ can be identified as an element of $\mathcal{R}$, denoted as $\tilde{f}^x$. In fact, $\tilde{f}^x_s = \delta_{\{f(\phi(x,s))\}}$ for each $s \in [0, \infty)$, whereas $x \in \mathbf{S} \to \tilde{f}^x \in \mathcal{R}$ is measurable. This measurable mapping $x \in \mathbf{S} \to \tilde{f}^x \in \mathcal{R}$ defines a deterministic stationary optimal strategy for the risk-sensitive DTMDP problem (B.3) by Proposition B.1. It is clear that the measurable mapping $x \in \mathbf{S} \to f(x) \in \mathbf{A}$ defines an optimal deterministic stationary policy for the PDMDP problem (5.3).

Finally, we show the remaining part of (a). Let H^* be a measurable $[1, \infty)$-valued function on $\mathbf{S}$ such that

$$\begin{aligned} & -(H^*(\phi(x,t)) - H^*(x)) \\ = \; & \int_0^t \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} H^*(y)\tilde{q}(dy|\phi(x,\tau), a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau), a))H^*(\phi(x,\tau)) \right\} d\tau, \\ & t \in [0, \infty), x \in \mathbf{S}. \end{aligned}$$

There exists a measurable mapping h from $\mathbf{S}$ to $\mathbf{A}$ such that

$$\begin{aligned}
&\inf_{a\in\mathbf{A}}\left\{\int_{\mathbf{S}} H^*(y)\tilde{q}(dy|x,a)-(q_x(a)-c(x,a))H^*(x)\right\}\\
=\ &\int_{\mathbf{S}} H^*(y)\tilde{q}(dy|x,h(x))-(q_x(h(x))-c(x,h(x)))H^*(x),\ \forall\ x\in\mathbf{S};
\end{aligned}$$

c.f., Proposition D.5 of [57]. It follows that $\int_0^s\int_{\mathbf{S}} H^*(y)\tilde{q}(dy|\phi(x,\tau),h(\phi(x,\tau)))d\tau$ is absolutely continuous in $s\in[0,t]$ for each $t\geq 0$. As in the proof of part (b),

$$\begin{aligned}
&\int_0^t e^{-\int_0^s (q_{\phi(x,v)}(h(\phi(x,v)))-c(\phi(x,v),h(\phi(x,v))))dv}\int_{\mathbf{S}} H^*(y)\tilde{q}(dy|\phi(x,s),h(\phi(x,s)))ds\\
&+e^{-\int_0^t (q_{\phi(x,s)}(h(\phi(x,s)))-c(\phi(x,s),h(\phi(x,s))))ds}H^*(\phi(x,t))-H^*(x)=0,\ \forall\ t\in[0,\infty),
\end{aligned}$$

and by passing to the lower limit as $t\to\infty$,

$$\begin{aligned}
H^*(x)\ \geq\ &\int_0^\infty e^{-\int_0^s (q_{\phi(x,v)}(h(\phi(x,v)))-c(\phi(x,v),h(\phi(x,v))))dv}\int_{\mathbf{S}} H^*(y)\tilde{q}(dy|\phi(x,s),h(\phi(x,s)))ds\\
&+e^{-\int_0^\infty q_{\phi(x,s)}(h(\phi(x,s)))ds}e^{\int_0^\infty c(\phi(x,s),h(\phi(x,s)))ds}\\
\geq\ &\inf_{\rho\in\mathcal{R}}\left\{\int_0^\infty e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds}\left(\int_{\mathbf{S}} H^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)\right)d\tau\right.\\
&\left.+e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds}e^{\int_0^\infty c(\phi(x,s),\rho_s)ds}\right\},\ \forall\ x\in\mathbf{S}. \qquad (5.11)
\end{aligned}$$

It remains to refer to Proposition B.1 for that $H^*(x)\geq V^*(x)$ for each $x\in\mathbf{S}$. □

Proof of Theorem 5.2. Let $V_0^*(x):=1$ for each $x\in\mathbf{S}$. For each $n\geq 0$, one can legitimately define

$$\begin{aligned}
V_{n+1}^*(x)\ =\ &\inf_{\rho\in\mathcal{R}}\left\{\int_0^\infty e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s)-c(\phi(x,s),\rho_s))ds}\left(\int_{\mathbf{S}} V_n^*(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau)\right)d\tau\right.\\
&\left.+e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds}e^{\int_0^\infty c(\phi(x,s),\rho_s)ds}\right\},\ \forall\ x\in\mathbf{S}. \qquad (5.12)
\end{aligned}$$

Recall that the DTMDP model $\{\mathbf{X},\mathbb{A},p,l\}$ satisfies Condition B.2, as noted in the proof of Lemma 5.1. Then by Proposition B.1, $\{V_n^*\}$ is a monotone nondecreasing sequence of $[1,\infty)$-valued measurable functions on $\mathbf{S}$ such that $V_n^*(x)\uparrow V^*(x)$ as

$n \uparrow \infty$, for each $x \in \mathbf{S}$.

Let $n \geq 0$ be fixed. As in Lemma 5.3, for each $x \in \mathbf{S}$, there is some $\rho^* \in \mathcal{R}$ such that

$$\begin{aligned} V_{n+1}^*(x) &= \inf_{\rho \in \mathcal{R}} \left\{ \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho_v) - c(\phi(x,v),\rho_v))dv} \int_{\mathbf{S}} V_n^*(y) \tilde{q}(dy|\phi(x,s),\rho_s) ds \right. \\ &\quad \left. + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} V_{n+1}^*(\phi(x,t)) \right\} \\ &= \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho_v^*) - c(\phi(x,v),\rho_v^*))dv} \int_{\mathbf{S}} V_n^*(y) \tilde{q}(dy|\phi(x,s),\rho_s^*) ds \\ &\quad + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s^*) - c(\phi(x,s),\rho_s^*))ds} V_{n+1}^*(\phi(x,t)), \ \forall\ t \geq 0. \end{aligned}$$

Also the relevant version of Lemma 5.2 holds: for each $x \in \mathbf{S}$ and $\rho \in \mathcal{R}$,

$$\begin{aligned} t \in [0,\infty) \ \rightarrow \ & \int_0^t e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \int_{\mathbf{S}} V_n^*(y) \tilde{q}(dy|\phi(x,\tau),\rho_\tau) d\tau \\ & + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} V_{n+1}^*(\phi(x,t)) \end{aligned}$$

is monotone nondecreasing in $t \in [0,\infty)$. Clearly, $V_{n+1}^*(\phi(x,t))$ is absolutely continuous in $t \in [0,\infty)$ for each $x \in \mathbf{S}$.

Corresponding to (5.9), we now have

$$\begin{aligned} 0 &= \int_0^t e^{-\int_0^s (q_{\phi(x,v)}(\rho_v^*) - c(\phi(x,v),\rho_v^*))dv} \int_{\mathbf{S}} V_n^*(y) \tilde{q}(dy|\phi(x,s),\rho_s^*) ds \\ &\quad + e^{-\int_0^t (q_{\phi(x,s)}(\rho_s^*) - c(\phi(x,s),\rho_s^*))ds} V_{n+1}^*(\phi(x,t)) - V_{n+1}^*(x) \\ &= \int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*) - c(\phi(x,v),\rho_v^*))dv} \left\{ \int_{\mathbf{S}} V_n^*(y) \tilde{q}(dy|\phi(x,\tau),\rho_\tau^*) + U_{n+1}^*(x,\tau) \right. \\ &\quad \left. - (q_{\phi(x,\tau)}(\rho_\tau^*) - c(\phi(x,\tau),\rho_\tau^*)) V_{n+1}^*(\phi(x,\tau)) \right\} d\tau \\ &\geq \int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*) - c(\phi(x,v),\rho_v^*))dv} \left\{ U_{n+1}^*(x,\tau) \right. \\ &\quad \left. + \inf_{a \in \mathbf{A}} \left\{ \int_{\mathbf{S}} V_n^*(y) \tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a)) V_{n+1}^*(\phi(x,\tau)) \right\} \right\} d\tau \\ &= \int_0^t e^{-\int_0^\tau (q_{\phi(x,v)}(\rho_v^*) - c(\phi(x,v),\rho_v^*))dv} \left\{ U_{n+1}^*(x,\tau) + \int_{\mathbf{S}} V_n^*(y) \tilde{q}(dy|\phi(x,\tau), f(\phi(x,\tau))) \right. \\ &\quad \left. - (q_{\phi(x,\tau)}(f(\phi(x,\tau))) - c(\phi(x,\tau), f(\phi(x,\tau)))) V_{n+1}^*(\phi(x,\tau)) \right\} d\tau, \end{aligned}$$

where $\tau \in [0,t] \to U^*_{n+1}(x,\tau)$ is integrable and coincides with $\frac{\partial V^*_{n+1}(\phi(x,t))}{\partial t}$ almost everywhere, and f is some measurable mapping from $\mathbf{S}$ to $\mathbf{A}$, whose existence is guaranteed by Proposition D.5 of [57]. Continued from the above relation, the reasoning in the proof of the first assertion in part (a) of Theorem 5.1 can be followed: eventually we see

$$U^*_{n+1}(x,\tau) + \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} V^*_n(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V^*_{n+1}(\phi(x,\tau)) \right\} = 0$$

almost everywhere in $\tau \in [0,t]$, i.e., the equation

$$\begin{aligned} & -(V(\phi(x,t)) - V(x)) \\ = & \int_0^t \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} V^*_n(y)\tilde{q}(dy|\phi(x,\tau),a) - (q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V(\phi(x,\tau)) \right\} d\tau, \\ & t \in [0,\infty), x \in \mathbf{S}, \end{aligned} \tag{5.13}$$

is satisfied by $V = V^*_{n+1}$.

Recall that $V^*_0 = V^{(0)}$. Suppose the recursive definition in (5.4) is valid up to step n, and $V^*_n(x) = V^{(n)}(x)$ for each $x \in \mathbf{S}$. Consider an arbitrarily fixed $[1,\infty)$-valued measurable solution V to (5.13), and let f^* be a measurable mapping from $\mathbf{S}$ to $\mathbf{A}$ such that

$$\begin{aligned} & \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} V^*_n(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))V(x) \right\} \\ = & \int_{\mathbf{S}} V^*_n(y)\tilde{q}(dy|x,f^*(x)) - (q_x(f^*(x)) - c(x,f^*(x)))V(x), \ \forall\ x \in \mathbf{S}. \end{aligned}$$

One can follow the reasoning in the last part of the proof of Theorem 5.1, and see, c.f. (5.11),

$$\begin{aligned} V(x) \ \geq \ & \int_0^\infty e^{-\int_0^s (q_{\phi(x,v)}(f^*(\phi(x,v))) - c(\phi(x,v),f^*(\phi(x,v))))dv} \int_{\mathbf{S}} V^*_n(y)\tilde{q}(dy|\phi(x,s),f^*(\phi(x,s)))ds \\ & + e^{-\int_0^\infty q_{\phi(x,s)}(f^*(\phi(x,s)))ds} e^{\int_0^\infty c(\phi(x,s),f^*(\phi(x,s)))ds} \\ \geq \ & \inf_{\rho\in\mathcal{R}} \left\{ \int_0^\infty e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \left(\int_{\mathbf{S}} V^*_n(y)\tilde{q}(dy|\phi(x,\tau),\rho_\tau) \right) d\tau \right. \\ & \left. + e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds} e^{\int_0^\infty c(\phi(x,s),\rho_s)ds} \right\} = V^*_{n+1}(x), \ \forall\ x \in \mathbf{S}, \end{aligned}$$

where the last equality is by (5.12). Thus, V^*_{n+1} is the minimal $[1,\infty)$-valued measurable solution to (5.13), and coincides with $V^{(n+1)}$. Therefore, by induction $V^*_n = V^{(n)}$ for each $n \geq 0$. It follows now that $V^{(n)}(x) \uparrow V^*(x)$ as $n \uparrow \infty$ for each $x \in \mathbf{S}$. $\square$

Part II

Other Problems on CTMDP and Stochastic Games

6 Discounted CTMDP with a lower bounding function

In this chapter, we consider the discounted CTMDP problems, where the negative part of each cost rate is bounded by a drift function, say w, whereas the positive part is allowed to be arbitrarily unbounded. Our focus is on the existence of a stationary optimal policy for the discounted CTMDP problems out of the more general class. Both constrained and unconstrained problems are considered. The investigations are based on the continuous-time version of the Veinott transformation. This technique was not widely employed in the previous literature in CTMDPs, but it clarifies the roles of the imposed conditions in a rather transparent way. As a consequence, we withdraw and weaken several conditions commonly imposed in the literature.

6.1 The constrained and unconstrained problems

For each $j = 0, 1, \ldots, N$, with $N \geq 1$ being a fixed integer, let c_j be a $(-\infty, \infty]$-valued measurable function on $\mathbb{K} = \{(x, a) | x \in \mathbf{S}, a \in \mathbf{A}(x)\}$, representing a cost rate, and d_j be a fixed finite constant, representing a corresponding constraint. We shall consider the following unconstrained and constrained α-discounted optimal control problems, respectively:

$$\text{Minimize over } \pi \in \Pi: \quad E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c_0(\xi_t, a) \pi(da|\omega, t) dt \right], \ x \in \mathbf{S}, \tag{6.1}$$

and

$$\begin{aligned} \text{Minimize over } \pi \in \Pi: \quad & E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c_0(\xi_t, a) \pi(da|\omega, t) dt \right] \\ \text{subject to} \quad & E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c_j(\xi_t, a) \pi(da|\omega, t) dt \right] \leq d_j, \; j = 1 \ldots N \end{aligned} \tag{6.2}$$

Here and below, we put

$$c(x_\infty, a) := 0, \; \forall \; a \in \mathbf{A} \bigcup \{a_\infty\}. \tag{6.3}$$

The conditions we impose below will ensure that the performance measures in the above two problems are well defined, though not necessarily finite.

A policy π is called *feasible* for the constrained problem (6.2) if it satisfies all the inequalities therein. A feasible policy π for problem (6.2) is said to be of a finite value if

$$E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c_0^{\pm}(\xi_t, a) \pi(da|\omega, t) dt \right] < \infty.$$

where $c_0^{\pm}$ denote the negative and positive part of function c_0.

A policy π^* is said to be optimal for problem (6.2) if it is feasible and satisfies

$$E_x^{\pi^*} \left[\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c_0(\xi_t, a) \pi^*(da|\omega, t) dt \right] \leq E_x^\pi \left[\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c_0(\xi_t, a) \pi(da|\omega, t) dt \right]$$

for each feasible policy π.

Note that the definition of optimality of a feasible policy for the constrained problem (6.2) requires a fixed initial state $x \in \mathbf{S}$. Here, we did not consider the more general case of a fixed initial distribution just for brevity and readability. The case of a fixed initial distribution γ can be similarly treated with additional conditions regarding γ.

We would like to allow the possibility of cost rates unbounded from both above and below. We consider the following set of conditions to guarantee that

the performance measures in problems (6.1) and (6.2) are well defined.

Condition 6.1. *There exists a* $[1,\infty)$*-valued measurable function* w *on* $\boldsymbol{S}$ *such that*

(a) for some finite constant $0 \leq \rho < \alpha$*,*

$$\int_{\boldsymbol{S}} w(y)q(dy|x,a) \leq \rho w(x),\ \forall\ (x,a) \in \mathbb{K};$$

(b) for some finite constant $L > 0$*,*

$$c_i^-(x,a) \leq Lw(x),\ \forall\ (x,a) \in \mathbb{K},\ i = 0,1,\ldots,N.$$

Here, for each $i = 0,1,\ldots,N$*,* c_i^- *is the negative part of the function* c_i*.*

Below, we allow that $w(x_\infty) := 0$. The cost rates satisfying part (b) of the above condition are said to be lower bounded by the drift function w; c.f. p.251 of [6] for a related definition for piecewise deterministic Markov decision processes.

Lemma 6.1. *Suppose Condition 6.1 is satisfied. Let a policy* π *be arbitrarily fixed. Then*

$$E_x^\pi\left[\int_0^\infty e^{-\alpha t}w(\xi_t)dt\right] < \infty,\ \forall\ x \in \boldsymbol{S}.$$

In particular, for each $x \in \boldsymbol{S}$*, the integrals* $E_x^\pi\left[\int_0^\infty e^{-\alpha t}\int_{\boldsymbol{A}} c_i(\xi_t,a)\pi(da|\omega,t)dt\right]$*,* $i = 0,1,\ldots,N$*, are well defined.*

Proof. This follows from Lemma 2 of [90] and (6.3). □

Assumption 1. *Throughout this paper, unless stated otherwise, Condition 6.1 is assumed to hold automatically, without specific reference.*

6.2 Conditions, statements and comments

Condition 6.2. *There exist a $(0,\infty)$-valued measurable function w' on $\boldsymbol{S}$ and a monotone nondecreasing sequence of measurable subsets $\{Z_m\}_{m=1}^{\infty} \subseteq \mathcal{B}(\boldsymbol{S})$ such that the following hold.*

(a) $Z_m \uparrow \boldsymbol{S}$ *as* $m \to \infty$.

(b) $\sup_{x\in Z_m} \overline{q}_x < \infty$ *for each* $m = 1, 2, \ldots$.

(c) For some constant $\rho' \in (0,\infty)$,

$$\int_{\boldsymbol{S}} w'(y) q(dy|x,a) \le \rho' w'(x), \ \forall\ x \in \boldsymbol{S},\ a \in \boldsymbol{A}(x).$$

(d) $\inf_{x\in \boldsymbol{S}\setminus Z_m} \frac{w'(x)}{w(x)} \to \infty$ *as* $m \to \infty$, *where the function* w *is from Condition 6.1.*

Condition 6.3. *(a) The multifunction* $x \in \boldsymbol{S} \mapsto \boldsymbol{A}(x) \in \mathcal{B}(\boldsymbol{A})$ *is compact-valued and upper semicontinuous.*

(b) For each w*-bounded continuous function* g *on* $\boldsymbol{S}$, $(x,a) \in \mathbb{K} \to \int_{\boldsymbol{S}} g(y)\tilde{q}(dy|x,a)$ *is continuous. Here and below the function* w *is from Condition 6.1.*

(c) The function w *is continuous on* $\boldsymbol{S}$, *and the functions* c_i *are lower semicontinuous on* $\mathbb{K}$.

The conditions formulated in the above can be satisfied when the negative part of each cost rate is bounded by a drift function, whereas the positive part is arbitrarily unbounded. In the literature of economics, such a cost rate might appear e.g., when one considers the logarithmic utility function, where they put $-\ln 0 := \infty$, see Section 7 of [97]; see also Example 2 of [69]. We formulate an example of such a CTMDP as follows.

Example 6.1. Consider a controlled M/M/∞ queueing system. The state $x \in \{0, 1, \ldots\} = \mathbf{S}$ represents the number of customers in the system. The control

is the arrival rate $a \in [0, x] \subseteq [0, \infty)$ for each $x \in \mathbf{S}$. The service rate $\mu > 0$ is uncontrolled. The cost rate is given by $c_0(x, a) = -\ln a$, and the constraint cost rate is given by $c_1(x, a) = x$. Then Conditions 6.1, 6.2 and 6.3 are all satisfied (for a large enough discount factor); one can put $w(x) = x+1$ and $w'(x) = 1+x^2$. On the other hand, there is no finite bounding function for $|c_0|$.

The next condition is for constrained problem only.

Condition 6.4. *There exists a feasible policy for problem (6.2) with a finite value.*

The main statement of this paper is the following one.

Theorem 6.1. *Suppose Conditions 6.1, 6.2 and 6.3 are satisfied. Then the following assertions hold.*

(a) There exists a deterministic stationary optimal policy for the unconstrained problem (6.1). In fact, one can always take a deterministic stationary policy providing the minimum in the equation (6.14) as a deterministic stationary optimal policy.

(b) If Condition 6.4 is also satisfied, then there exists a stationary optimal policy for the constrained problem (6.2).

In the previous literature, general discounted CTMDPs have not been considered when the cost rates were bounded below by a lower bounding function, and arbitrarily unbounded from the above, although for specific piecewise deterministic Markov decision processes with jumps driven by a Poisson process, this was considered in [6] following a different method. Discrete-time problems with a lower bounding function were considered in [6, 68], and in latter reference, the motivation for considering such cost functions was explained with their applications to economics. For discounted DTMDP problems, the treatment in [6, 68] was direct. But it is possible to reduce this to equivalent problems with nonnegative cost functions, using the technique in p.101 of [99], see also [29] and p.79 of

[2]. The proof of Theorem 6.1 will be based on a similar technique for CTMDPs, which, to the best of our knowledge, has not been widely applied to CTMDPs.

For the more restrictive case, where the cost rates are w-bounded, with w coming from Condition 6.1, Theorem 6.1(a) was obtained in [10] under essentially equivalent conditions for discounted CTMDPs in a denumerable state space but restricted to the class of stationary policies. Here we show that it is without loss of generality to be restricted to this narrower class of policies under the imposed conditions. Otherwise, this sufficiency result seems not to follow from other known results in the relevant literature. The approach in [10] was directly based on the application of the Dynkin's forumla, and is different from ours. When the cost rates are only lower w-bounded, the value function is in general not w-bounded. Since under the conditions in [10] and here, Dynkin's formula is only applicable to the class of w-bounded functions, the treatment in [10] does not directly apply to the general case dealt with here.

Also when the cost rates are w-bounded, Theorem 6.1(b) was obtained in e.g., [89] but under stronger conditions. We include them here for ease of reference.

Instead of Condition 6.2, the following condition was imposed in [89].

Condition 6.5. *There exists a $(0,\infty)$-valued measurable function $\widetilde{w}'$ on $\boldsymbol{S}$ such that the following hold.*

(a) For some constant $\widetilde{L}' \in (0,\infty)$, $\overline{q}_x \le \widetilde{L}'\widetilde{w}'(x)$ for each $x \in \boldsymbol{S}$.

(b) For some constant $\widetilde{\rho}' \in (0,\infty)$, $\int_{\boldsymbol{S}} \widetilde{w}'(y) q(dy|x,a) \le \widetilde{\rho}'\widetilde{w}'(x)$ for each $(x,a) \in \mathbb{K}$.

(c) For some constant $\widetilde{L} \in (0,\infty)$, $(\overline{q}_x+1)w(x) \le \widetilde{L}\widetilde{w}'(x)$ for each $x \in \boldsymbol{S}$, where the function w comes from Condition 6.1.

It is easy to see that, if the above condition is satisfied, then so is Condition 6.2 with $w' = \widetilde{w}'+1$, $\rho' = \widetilde{\rho}'$, $Z_m = \left\{x \in \mathbf{S} : \ \frac{\widetilde{w}'(x)+1}{w(x)} \le m\right\}$ for each $m = 1, 2, \ldots$.

Furthermore, under Conditions 6.1, 6.2 and 6.4, in addition to Condition 6.3, it was also assumed in [89] that the function $\frac{\widetilde{w}'}{w}$ is a moment function on

$\mathbb{K}$, see Definition E.7 of [57], in order to apply the Prokhorov theorem in their proof, see Proposition E.8 and Theorem E.6 of [57]. This is not needed here. The investigations in [89] are largely based on the Dynkin's formula, and do not handle the more general cost rates considered here.

The rest of this section proves Theorem 6.1. On the way, we comment and clarify the roles of the imposed conditions, and present the auxiliary statements.

6.3 Proof of the main statement

The proof of Theorem 6.1 follows from a sequence of lemmas. The outline of the proof steps is announced in the next remark.

Remark 6.1. The main themes in the proof of Theorem 6.1 can be summarized as follows.

1. Under Condition 6.1, the w-transformation, see Lemma 6.3, allows one to reduce the original problems (6.1) and (6.2) to problems (6.5) and (6.6) for the w-transformed CTMDP model with cost rates bounded from below, equivalently.

2. Under the extra Condition 6.2, problems (6.5) and (6.6) are reduced to discounted CTMDP problems (6.8) and (6.9) with nonnegative cost rates by adding some large enough constant. This is possible because Condition 6.2 ensures that the controlled process in the w-transformed CTMDP model is nonexplosive under each Markov policy, according to Lemma 6.4.

3. By applying the reduction technique in [33, 35], discounted CTMDP problems (6.8) and (6.9) with nonnegative cost rates are reduced to total undiscounted DTMDP problems (6.12) and (6.13) with nonnegative cost functions.

4. Apply the optimality results in [26] to the DTMDP problems (6.12) and

(6.13) with nonnegative cost functions. Then deduce from here the corresponding optimality results for the original problems (6.1) and (6.2).

The details are as follows.

Proof of Theorem 6.1. The following statement is a consequence of Theorem 4.2 of [37], see also [36], and is the starting point of our reasoning.

Lemma 6.2. *For each initial state $x \in \boldsymbol{S}$ and policy π, there exists a Markov policy φ such that*

$$E_x^{\pi}\left[\int_0^{\infty} e^{-\alpha t}\int_{\boldsymbol{A}} f(\xi_t,a)\pi(da|\omega,t)dt\right] = E_x^{\varphi}\left[\int_0^{\infty} e^{-\alpha t}\int_{\boldsymbol{A}} f(\xi_t,a)\varphi(da|\xi_t,t)dt\right]$$

for each $[0,\infty]$-valued measurable function f on $\mathbb{K}$.

The above lemma implies that without loss of generality, we can restrict to the class of Markov policies for problems (6.1) and (6.2), i.e., if we obtain an optimal policy out of the class of Markov policies for problem (6.1) (or (6.2)), then that policy is optimal for problem (6.1) (or (6.2)) out of the general class.

We recall some definitions related to the process $\{\xi_t, t \geq 0\}$ under a Markov policy φ. Let us consider the signed kernel on $\mathbf{S}$ from $\mathbf{S}\times[0,\infty)$ defined by

$$q_{\varphi}(dy|x,t) := \int_{\mathbf{A}} q(dy|x,a)\varphi(da|x,t),\ \forall\ x \in \mathbf{S},\ t \in [0,\infty).$$

Then q_{φ} is a conservative and stable Q-function in the sense of [38, p.262]. For the ease of reference, we recall some relevant definitions and facts about Q-functions in the appendix.

According to Theorem 2.2 of [38], under a Markov policy, say φ, the process $\{\xi_t, t \geq 0\}$ is a Markov pure jump process on $\{\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P^{\varphi}\}$, that is, for each $s,t \in [0,\infty)$,

$$P^{\varphi}(\xi_{t+s} \in \Gamma|\mathcal{F}_t) = P^{\varphi}(\xi_{t+s} \in \Gamma|\xi_t),\ \forall\ \Gamma \in \mathcal{B}(\mathbf{X}_{\infty});$$

and each trajectory of $\{\xi_t; t \geq 0\}$ is piecewise constant and right-continuous, such that for each $t \in [0, t_\infty)$, there are finitely many discontinuity points on the interval $[0, t]$, see Definition 1 in Chapter III of [44]. Here and below, we omit the subscript in P^φ_γ, whenever the initial distribution γ is irrelevant. Furthermore, by Theorem 2.2 of [38], p_{q_φ} defined by (A.1) with q being replaced by q_φ is the transition function corresponding to the process $\{\xi_t, t \geq 0\}$, i.e., for each $s \leq t$, on $\{s < t_\infty\}$,

$$P^\varphi(\xi_t \in \Gamma | \mathcal{F}_s) = p_{q_\varphi}(s, \xi_s, t, \Gamma), \ \forall\ \Gamma \in \mathcal{B}(\mathbf{S}),$$

c.f. p.1397 of [76]. Consequently, for each Markov policy φ,

$$E^\varphi_x \left[\int_0^\infty e^{-\alpha t} \int_{\mathbf{A}} c_i(\xi_t, a) \varphi(da|\xi_t, t) dt \right] = \int_0^\infty \int_{\mathbf{S}} e^{-\alpha t} \int_{\mathbf{A}} c_i(y, a) \varphi(da|y, t) p_{q_\varphi}(0, x, t, dy) dt$$

for each $i = 0, 1, \ldots, N$ and $\forall\ x \in \mathbf{S}$.

Given the Q-function q_φ on S induced by a Markov policy φ, let us introduce the w-transformed Q-function q^w_φ on $\mathbf{S}_\delta$ defined as follows.

Let

$$\mathbf{S}_\delta := \mathbf{S} \bigcup \{\delta\}$$

with $\delta \notin \mathbf{S}$ being an isolated point concerning the topology of $\mathbf{S}_\delta$ that satisfies $\delta \neq x_\infty$. The w-transformed (stable conservative) Q-function q^w_φ on $\mathbf{S}_\delta$ is defined by

$$q^w_\varphi(\Gamma|x, s) := \begin{cases} \frac{\int_\Gamma w(y) q_\varphi(dy|x,s)}{w(x)}, & \text{if } x \in \mathbf{S},\ \Gamma \in \mathcal{B}(\mathbf{S}),\ x \notin \Gamma; \\ \rho - \frac{\int_S w(y) q_\varphi(dy|x,s)}{w(x)}, & \text{if } x \in \mathbf{S},\ \Gamma = \{\delta\}; \\ 0, & \text{if } x = \delta,\ \Gamma = \mathbf{S}_\delta. \end{cases} \tag{6.4}$$

for each $s \in [0, \infty)$; and

$$q^w_{\varphi_x}(s) := \rho + q_{\varphi_x}(s), \ \forall \ s \in [0, \infty).$$

Here, $q_{\varphi_x}(s) = -q_\varphi(\mathbf{S} \setminus \{x\}|x, s)$; see the appendix for more definitions and relevant notations concerning a Q-function. This transformation is the continuous-time version of the Veinott transformation, see [100], widely known in the literature of DTMDPs. For (uncontrolled) homogeneous continuous-time Markov chains, this transformation was used in e.g., [3, 96, 95].

Lemma 6.3. *Let a Markov policy φ be fixed. For each $x \in \boldsymbol{S}$, $s, t \in [0, \infty)$, $s \le t$ and $\Gamma \in \mathcal{B}(\boldsymbol{S})$, the following relation holds;*

$$p_{q^w_\varphi}(s, x, t, \Gamma) = \frac{e^{-\rho(t-s)}}{w(x)} \int_\Gamma w(y) p_{q_\varphi}(s, x, t, dy).$$

Proof. See Lemma A.3 of [107]. □

By Lemma 6.3, we see that for each $i = 0, 1, \ldots, N$,

$$\begin{aligned} & w(x) \int_0^\infty \int_{\mathbf{S}} p_{q^w_\varphi}(0, x, t, dy) \int_{\mathbf{A}} \frac{c_i(y, a)}{w(y)} \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \\ = & \int_0^\infty \int_{\mathbf{S}} \int_{\mathbf{A}} c_i(y, a) \varphi(da|y, t) e^{-\alpha t} p_{q_\varphi}(0, x, t, dy) dt, \ \forall \ x \in \mathbf{S}. \end{aligned}$$

Hence, problem (6.1) is equivalent to

$$\text{Minimize over } \varphi \in \Pi^M\text{:} \quad \int_0^\infty \int_{\mathbf{S}} p_{q^w_\varphi}(0, x, t, dy) \int_{\mathbf{A}} \frac{c_0(y, a)}{w(y)} \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \tag{6.5}$$

and problem (6.2) is equivalent to

$$\begin{aligned} \text{Minimize over } \varphi \in \Pi^M\text{:} \quad & \int_0^\infty \int_{\mathbf{S}} p_{q^w_\varphi}(0, x, t, dy) \int_{\mathbf{A}} \frac{c_i(y, a)}{w(y)} \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \\ \text{subject to} \quad & \int_0^\infty \int_{\mathbf{S}} p_{q^w_\varphi}(0, x, t, dy) \int_{\mathbf{A}} \frac{c_j(y, a)}{w(y)} \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \le \frac{d_j}{w(x)}, \\ & j = 1, 2, \ldots, N. \end{aligned} \tag{6.6}$$

Thus, one can consider the w-transformed CTMDP $\{\mathbf{S}_\delta, \mathbf{A}\bigcup\{a_\infty\}, \mathbf{A}_\delta(\cdot), q^w\}$, where $\mathbf{A}_\delta(\delta) := \{a_\infty\}$, and $\mathbf{A}_\delta(x) := \mathbf{A}(x)$ for each $x \in \mathbf{S}$, while the transition rate q^w is defined by, c.f. (6.4),

$$q^w(\Gamma|x,a) = \begin{cases} \frac{\int_\Gamma w(y)q(dy|x,a)}{w(x)}, & \text{if } x \in \mathbf{S},\ \Gamma \in \mathcal{B}(\mathbf{S}),\ x \notin \Gamma; \\ \rho - \frac{\int_S w(y)q(dy|x,a)}{w(x)}, & \text{if } x \in \mathbf{S},\ \Gamma = \{\delta\}; \\ 0, & \text{if } x = \delta,\ \Gamma = \mathbf{S}_\delta. \end{cases}$$

for each $x \in \mathbf{S}_\delta$ and $a \in \mathbf{A}_\delta(x)$; and

$$q_x^w(a) := \rho + q_x(a),\ \forall\ x \in \mathbf{S},\ a \in \mathbf{A}_\delta(x).$$

The requirement of $\alpha > \rho$ in Condition 6.1(a) is needed so that problems (6.5) and (6.6) are legitimate $(\alpha - \rho)$-discounted problems of the w-transformed CTMDP with the cost rates c_i^w defined by

$$c_i^w(x,a) := \frac{c_i(x,a)}{w(x)}$$

for each $x \in \mathbf{S}$, $a \in \mathbf{A}(x)$; and

$$c_i^w(\delta, a_\infty) := 0.$$

According to the reduction technique for discounted CTMDPs, see [35], the CTMDP problems (6.5) and (6.6) can be reduced to equivalent total undiscounted problems for the DTMDP $\{\mathbf{S}_\delta\bigcup\{x_\infty\}, \mathbf{A}\bigcup\{a_\infty\}, \mathbf{A}_\delta(\cdot), T\}$ with the cost functions C_i, where the transition probability T is defined by

$$T(\Gamma|x,a) := \frac{\int_\Gamma w(y)q(dy|x,a)}{(\alpha + q_x(a))w(x)}$$

for each $\Gamma \in \mathcal{B}(\mathbf{S})$, $x \notin \Gamma$, and $a \in \mathbf{A}_\delta(x)$;

$$T(\{\delta\}|x,a) := \frac{\rho w(x) - \int_{\mathbf{S}} w(y) q(dy|x,a)}{(\alpha + q_x(a))w(x)}$$

for each $x \in \mathbf{S}$ and $a \in \mathbf{A}_\delta(x)$;

$$T(\{x_\infty\}|x,a) := \frac{\alpha - \rho}{\alpha + q_x(a)}$$

for each $x \in \mathbf{S}$ and $a \in \mathbf{A}_\delta(x)$; and $T(\{x_\infty\}|x_\infty, a_\infty) := 1 =: T(\{x_\infty\}|\delta, a_\infty)$, and the cost functions C_i are defined by

$$C_i(x,a) := \frac{c_i(x,a)}{(\alpha + q_x(a))w(x)}$$

for each $x \in \mathbf{S}$ and $a \in \mathbf{A}_\delta(x)$; and

$$C_i(\delta, a_\infty) := 0 =: C_i(x_\infty, a_\infty).$$

More precisely, given the initial state $x \in \mathbf{S}$, for each Markov policy φ for the w-transformed CTMDP, there is a strategy σ for the DTMDP $\{\mathbf{S}_\delta \bigcup \{x_\infty\}, \mathbf{A} \bigcup \{a_\infty\}, \mathbf{A}_\delta(\cdot), T\}$ such that

$$\int_0^\infty \int_{\mathbf{S}} p_{q_\varphi^w}(0,x,t,dy) \frac{c_i(y,a)}{w(y)} e^{-(\alpha-\rho)t} dt = \mathbb{E}_x^\sigma \left[\sum_{n=0}^\infty C_i(X_n, A_n) \right]$$

for each $i = 0, 1, \ldots, N$, and vice versa. Moreover, in the previous equality, if φ is a deterministic stationary (respectively, stationary) policy, then σ can be taken as a deterministic stationary (respectively, stationary) strategy for the DTMDP, and vice versa. Here $\{\mathbf{X}_n\}$ and $\{\mathbf{A}_n\}$ are the controlled and controlling processes in the DTMDP. The term "strategy" is reserved for the DTMDP to avoid the potential confusion with the corresponding notion for the CTMDP. We refer the reader to e.g., [57, 86] for the standard description of a DTMDP.

Note that in general, the DTMDP $\{\mathbf{S}_\delta \bigcup \{x_\infty\}, \mathbf{A} \bigcup \{a_\infty\}, \mathbf{A}_\delta(\cdot), T\}$ is not ab-

sorbing in the sense of [2, 39], and the cost function C_i can take both positive and negative values. We formulate such a CTMDP in the next example.

Example 6.2. Suppose the CTMDP is an uncontrolled pure birth process with $\mathbf{S} = \{1, 2, \dots\}$. The birth rate at the state $x \in \mathbf{S}$ is $2x$. The discount factor is $\alpha = 2$. We put $\rho = 0$ and $w(x) = 1$ for each $x \in \mathbf{S}$. Suppose the cost rate is only zero at the state δ. For the induced DTMDP, $\{x_\infty\}$ is the absorbing set; the point δ can be excluded from the state space because it is never reached starting from $\mathbf{S} \bigcup \{x_\infty\}$. Then one can show that starting from 1, the expected time until the DTMDP reaches x_∞ is infinite. In accordance with e.g., [2, 39], this means that the model is not absorbing, i.e., the expected time to absorption is not finite.

On the other hand, the functions c_i^w, $i = 0, 1, \dots, N$, are bounded from below under Condition 6.1(b). Let some common lower bound be $\underline{c} \leq 0$. Let

$$\widetilde{c}_i^w := c_i^w - \underline{c} \tag{6.7}$$

for each $i = 0, 1, \dots, N$. Then the functions $\widetilde{c}_i^w$ are all nonnegative. In order for problems (6.5) and (6.6) to be equivalent to

$$\text{Minimize over } \varphi \in \Pi^M\text{:} \qquad \int_0^\infty \int_{\mathbf{S}_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{\mathbf{A}_\delta} \widetilde{c}_0^w(y, a) \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \tag{6.8}$$

and

$$\begin{aligned} \text{Minimize over } \varphi \in \Pi^M\text{:} \quad & \int_0^\infty \int_{\mathbf{S}_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{\mathbf{A}_\delta} \widetilde{c}_0^w(y) \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \\ \text{such that} \quad & \int_0^\infty \int_{\mathbf{S}_\delta} p_{q_\varphi^w}(0, x, t, dy) \int_{\mathbf{A}_\delta} \widetilde{c}_j^w(y) \varphi(da|y, t) e^{-(\alpha-\rho)t} dt \leq \frac{d_j}{w(x)} - \frac{\underline{c}}{\alpha - \rho}, \\ & j = 1, 2, \dots, N, \end{aligned} \tag{6.9}$$

respectively, we need the following relation to hold for each $\varphi \in \Pi^M$:

$$p_{q_\varphi^w}(0, x, t, \mathbf{S}_\delta) = 1, \ \forall\ x \in \mathbf{S}, \ t \in [0, \infty). \tag{6.10}$$

In general, problems (6.5) and (6.6) are not equivalent to problems (6.8) and

(6.9). We demonstrate this with the following example, which was also considered by Spieksma in [95].

Example 6.3. Let $\mathbf{S} = \{0, 1, 2, \dots\}$, $\mathbf{A}(x) \equiv \mathbf{A} = \{0, 1\}$. We endow them with the discrete topology. The transition rate is given by

$$q(\{y\}|x, 0) = \begin{cases} \frac{5}{12}2^x, & \text{if } x \neq 0,\ y = x + 1; \\ \frac{7}{12}2^x, & \text{if } x \neq 0,\ y = x - 1; \\ 0, & \text{if } x = 0. \end{cases}$$

and $q(\{y\}|x, 1) = 0$ for each $x, y \in \mathbf{S}$. Let $w(x) = \left(\frac{7}{5}\right)^x$ for each $x \in \mathbf{S}$. Then one can verify that

$$\sum_{y \in \mathbf{S}} w(y) q(\{y\}|x, a) = 0,\ \forall\ x \in \mathbf{S},\ a \in \mathbf{A},$$

and so let $\rho = 0$, and $\alpha = 1$. Let $c_0(x, a) \equiv 0$. Put $\underline{c} = -1$. Conditions 6.1 and 6.3 are satisfied.

Now

$$q^w(\{y\}|x, 0) = \begin{cases} \frac{7}{12}2^x, & \text{if } x \neq \delta,\ x \neq 0,\ y = x + 1; \\ \frac{5}{12}2^x, & \text{if } x \neq \delta,\ x \neq 0,\ y = x - 1; \\ 0, & \text{if } x \neq \delta,\ y = \delta; \\ 0, & \text{if } x = \delta \text{ or } x = 0. \end{cases}$$

and $q^w_x(0) = 2^x$ for each $x \neq \delta,\ 0$, and $q^w_x(0) = 0$ if $x = 0,\ \delta$. Also $q^w_x(1) = 0$ for each $x \in \mathbf{S}_\delta$.

Consider the following two deterministic stationary strategies: $\varphi_0(da|x, t) \equiv \delta_0(da)$ and $\varphi_1(da|x, t) \equiv \delta_1(da)$. Clearly, they are both optimal for problem (6.5). On the other hand,

$$\int_0^\infty \int_{\mathbf{S}_\delta} p_{q^w_{\varphi_i}}(0, x, t, dy) \int_{\mathbf{A}_\delta} \widetilde{c}^w_0(y, a) \varphi_i(da|y, t) e^{-(\alpha-\rho)t} dt = \int_0^\infty p_{q^w_{\varphi_i}}(0, x, t, \mathbf{S}_\delta) e^{-t} dt$$
$x \in \mathbf{S},\ i = 0, 1.$

Clearly, $p_{q^w_{\varphi_1}}(0,x,t,\mathbf{S}_\delta) \equiv 1 = \int_0^\infty p_{q^w_{\varphi_1}}(0,x,t,\mathbf{S}_\delta)e^{-t}dt$. It is shown in Section 5 of [95] that (6.10) does not hold for $\varphi = \varphi_0$ with some $x \in \mathbf{S}$; this can also be checked using Theorem 2 of [14]. It follows that for some $x \in \mathbf{S}$, $\int_0^\infty p_{q^w_{\varphi_0}}(0,x,t,\mathbf{S}_\delta)e^{-t}dt < 1$; see also Lemma 2.1 of [107]. Therefore, the policy φ_1 is not optimal for problem (6.8), although it is optimal for problem (6.5). Hence, in general, (6.5) and (6.6) are not equivalent to problems (6.8) and (6.9).

Remark 6.2. Example 6.3 illustrates the role of the requirement (6.10). Condition 6.2 is precisely imposed for this purpose, as seen in the next statement. (An alternative justification of the role of Condition 6.2 is that it validates the Dynkin's formula for the original CTMDP to a certain class of functions, see [10] for the homogeneous denumerable case. But the explanation here is more transparent in our opinion.) In the literature, e.g., [48, 89, 91], stronger conditions, e.g., Condition 6.5, than Condition 6.2, were imposed to guarantee (6.10) to hold. The investigations there were not based on reduction method to DTMDP.

Lemma 6.4. *Let some Markov policy φ be fixed. Suppose Condition 6.1(a) and Condition 6.2 are satisfied. Then (6.10) holds.*

Proof. According to Theorem A.1, for the statement it suffices to verify that Condition A.1 is satisfied.

Since the Markov policy φ is fixed throughout this proof, we write q_φ as q for brevity. Note that for $\forall x \in \mathbf{S}, s \geq 0$

$$\begin{aligned} \int_{\mathbf{S}} \frac{w'(y)}{w(y)} q^w(dy|x,s) &= \int_{\mathbf{S}} \frac{w'(y)}{w(y)} \frac{w(y)}{w(x)} \widetilde{q}(dy|x,s) - (\rho + q_x(s)) \frac{w'(x)}{w(x)} \\ = \int_{\mathbf{S}} \frac{w'(y)}{w(x)} \widetilde{q}(dy|x,s) - (\rho + q_x(s)) \frac{w'(x)}{w(x)} &\leq (\rho' - \rho) \frac{w'(x)}{w(x)} \end{aligned} \tag{6.11}$$

Consider the $[0,\infty)$-valued measurable function $\widetilde{w}$ on $[0,\infty) \times \mathbf{S}_\delta$ defined for each $v \in [0,\infty)$ by $\widetilde{w}(v,x) = \frac{w'(x)}{w(x)}$ if $x \in \mathbf{S}$ and $\widetilde{w}(v,\delta) = 0$. Then Condition A.1, with $\mathbf{S}$ and q being replaced by $\mathbf{S}_\delta$ and q^w, is satisfied by the monotone nondecreasing sequence of measurable subsets $\{\widetilde{V}_n\}_{n=1}^\infty$ of $\mathbb{R}_+^0 \times \mathbf{S}_\delta$ defined by

$\widetilde{V}_n = [0,\infty) \times V_n \bigcup \{\delta\}$ for each $n = 1, 2, \ldots$, and the function $\widetilde{w}$ on $[0,\infty) \times \mathbf{S}_\delta$ defined in the above. In greater detail, part (d) of the corresponding version of Condition A.1 is satisfied because, by (6.11),

$$\begin{aligned} & \int_0^\infty \int_{\mathbf{S}_\delta} \widetilde{w}(t+v,y) e^{-\rho' t - \int_{(0,t]} q_x^w(s+v)ds} \widetilde{q^w}(dy|x,t+v)dt \\ \leq & \int_0^\infty e^{-\rho' t - \int_0^t q_x^w(s+v)ds} \left(q_x(s)+\rho'\right) \widetilde{w}(v,x)dt = \widetilde{w}(v,x), \ \forall\ x \in \mathbf{S}, \end{aligned}$$

and the last inequality holds trivially when $x = \delta$.

Thus, by Theorem A.1, we see that relation (6.10) is satisfied, and the statement follows. □

By the way, under Condition 6.1(a), in certain models, Condition 6.2 is also necessary for (6.10) to hold under certain policies; see [107]. In the homogeneous denumerable case, this was first observed in [96]. For more concrete examples such as single birth processes, this necessity part was known earlier, see [15].

As a result of the above lemma and the discussions above it, we see that under Condition 6.1 and Condition 6.2, one can reduce the α-discounted problems (6.1) and (6.2) for the original CTMDP $\{\mathbf{S}, \mathbf{A}, \mathbf{A}(\cdot), q\}$ to the $(\alpha-\rho)$-discounted problems (6.8) and (6.9) for the CTMDP $\{\mathbf{S}_\delta, \mathbf{A}_\delta, \mathbf{A}_\delta(\cdot), q^w\}$ with nonnegative cost rates. Furthermore, according to the reduction technique [35], which was also sketched in the above, problems (6.8) and (6.9) can be reduced to

$$\text{Minimize over } \sigma \qquad \mathbb{E}_x^\sigma \left[\sum_{n=0}^\infty \widetilde{C}_0(\mathbf{X}_n, \mathbf{A}_n) \right], \ x \in \mathbf{S}, \tag{6.12}$$

and

$$\begin{aligned} \text{Minimize over } \sigma: \quad & \mathbb{E}_x^\sigma \left[\sum_{n=0}^{\infty} \widetilde{C}_0(\mathbf{X}_n, \mathbf{A}_n) \right] \\ \text{such that} \quad & \mathbb{E}_x^\sigma \left[\sum_{n=0}^{\infty} \widetilde{C}_j(\mathbf{X}_n, \mathbf{A}_n) \right] \le \frac{d_j}{w(x)} - \frac{\underline{c}}{\alpha - \rho}, \\ & j = 1, 2, \ldots, N, \end{aligned} \tag{6.13}$$

respectively, for the DTMDP $\{\mathbf{S}_\delta \bigcup \{x_\infty\}, \mathbf{A} \bigcup \{a_\infty\}, \mathbf{A}_\delta(\cdot), T\}$ defined earlier. Here the cost functions $\widetilde{C}_i$ for the DTMDP are defined by

$$\widetilde{C}_i(x, a) := \frac{\widetilde{c}_i^w(x, a)}{(\alpha + q_x(a))} \ge 0$$

for each $x \in \mathbf{S}_\delta$ and $a \in \mathbf{A}_\delta(x)$; and

$$\widetilde{C}_i(x_\infty, a_\infty) := 0,$$

with the functions $\widetilde{c}_i^w$ being defined by (6.7). Note that the cost functions $\widetilde{C}_i$ could be arbitrarily unbounded from above.

Finally, if Condition 6.1, Condition 6.2, and Condition 6.3 are all satisfied, then it is easy to check that the DTMDP $\{\mathbf{S}_\delta \bigcup \{x_\infty\}, \mathbf{A} \bigcup \{a_\infty\}, \mathbf{A}_\delta(\cdot), T\}$ with the nonnegative cost functions $\widetilde{C}_i$ is a semicontinuous model, see [6, 30], and it is a standard result that there exists an optimal deterministic stationary strategy for problem (6.12). For the constrained problem (6.13), under the extra Condition 6.4, one can refer to Theorem 4.1 of [26], see also Theorem A.2 of [20], for the existence of a stationary optimal strategy for (6.13). Since these two DTMDP problems are equivalent to the original CTMDP problems, according to the reduction technique for discounted CTMDP problems as mentioned earlier, we immediately conclude the existence of an optimal deterministic stationary policy for the unconstrained CTMDP problem (6.1) and an optimal stationary policy for the constrained CTMDP problem (6.2). The proof of Theorem 6.1 is thus

completed. □

We finish this section with the following observation. Suppose Conditions 6.1 and 6.3 are satisfied. If one solves problem (6.8) with a deterministic stationary policy φ, which also satisfies (6.10), then φ is also optimal for problem (6.5), in spite that Condition 6.2 has not been assumed to hold uniformly in all actions.

The justifications are this claim are as follows. In general, problems (6.5) and (6.6) are not equivalent to (6.8) and (6.9), respectively; recall Example 6.3. According to [35], (6.8) is equivalent to the DTMDP problem $\{\mathbf{S}_\delta \bigcup \{x_\infty\}, \mathbf{A} \bigcup \{a_\infty\}, \mathbf{A}_\delta(\cdot), T\}$ with the cost function $\widetilde{C}_0$. Suppose φ^* is an optimal deterministic strategy for this DTMDP problem. Under Conditions 6.1 and Condition 6.3, if W_α^* denotes the value function of this DTMDP problem, then such an optimal deterministic stationary strategy exists and can be obtained by taking the measurable selector providing the minimum in the following:

$$W_\alpha^*(x) = \inf_{a \in \mathbf{A}_\delta(x)} \left\{ \widetilde{C}_0(x,a) + \int_{\mathbf{S}_\delta} T(dy|x,a) V^*(y) \right\}, \ \forall \ x \in \mathbf{S}_\delta. \tag{6.14}$$

We claim that φ^* is also an optimal deterministic policy for the CTMDP problem (6.5), provided that (6.10) holds for this particular strategy φ^*, i.e.,

$$p_{q_{\varphi^*}^w}(0,x,t,\mathbf{S}_\delta) = 1, \ \forall \ x \in \mathbf{S}, \ t \in [0,\infty). \tag{6.15}$$

Indeed, since φ^* is optimal for the DTMDP $\{\mathbf{S}_\delta \bigcup \{x_\infty\}, \mathbf{A} \bigcup \{a_\infty\}, \mathbf{A}_\delta(\cdot), T\}$ with the cost function $\widetilde{C}_0$, which is equivalent to problem (6.8),

$$\begin{aligned}
& \inf_{\varphi \in \Pi^M} \left\{ \int_0^\infty \int_{\mathbf{S}_\delta} p_{q_\varphi^w}(0,x,t,dy) \int_{\mathbf{A}_\delta} \widetilde{c}_0^w(y,a) \varphi(da|y,t) e^{-(\alpha-\rho)t} dt \right\} \\
= & \int_0^\infty \int_{\mathbf{S}_\delta} p_{q_{\varphi^*}^w}(0,x,t,dy) \widetilde{c}_0^w(y,\varphi^*(y)) e^{-(\alpha-\rho)t} dt \\
= & \int_0^\infty \int_{\mathbf{S}} p_{q_{\varphi^*}^w}(0,x,t,dy) \frac{c_0(y,\varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt - \frac{\underline{c}}{\alpha-\rho}, \ \forall \ x \in \mathbf{S}.
\end{aligned}$$

Consider an arbitrarily fixed $\varphi \in \Pi^M$. Then for each $x \in \mathbf{S}$,

$$\begin{aligned}
&\int_0^\infty \int_{\mathbf{S}} p_{q^w_{\varphi^*}}(0,x,t,dy) \frac{c_0(y,\varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt - \frac{\underline{c}}{\alpha-\rho} \\
\leq\ & \int_0^\infty \int_{\mathbf{S}_\delta} p_{q^w_{\varphi}}(0,x,t,dy) \int_{\mathbf{A}_\delta} \widetilde{c}^w_0(y,a) \varphi(da|y,t) e^{-(\alpha-\rho)t} dt \\
=\ & \int_0^\infty \int_{\mathbf{S}} p_{q^w_{\varphi}}(0,x,t,dy) \int_{\mathbf{A}} \frac{c_0(y,a)}{w(y)} \varphi(da|y,t) e^{-(\alpha-\rho)t} dt - \underline{c} \int_0^\infty p_{q^w_{\varphi}}(0,x,t,\mathbf{S}_\delta) e^{-(\alpha-\rho)t} dt.
\end{aligned}$$

Since $\underline{c} \leq 0$, and $p_{q^w_{\varphi}}(0,x,t,\mathbf{S}_\delta) \leq 1$, it follows that

$$\begin{aligned}
&\int_0^\infty \int_{\mathbf{S}} p_{q^w_{\varphi^*}}(0,x,t,dy) \frac{c_0(y,\varphi^*(y))}{w(y)} e^{-(\alpha-\rho)t} dt \\
\leq\ & \int_0^\infty \int_{\mathbf{S}} p_{q^w_{\varphi}}(0,x,t,dy) \int_{\mathbf{A}} \frac{c_0(y,a)}{w(y)} \varphi(da|y,t) e^{-(\alpha-\rho)t} dt,\ \forall\ x \in \mathbf{S}.
\end{aligned}$$

Condition (6.15) can be checked using Theorem A.1 in the appendix. The similar reasoning also holds for the constrained problem. To avoid repetition, we omit the details.

7 Zero-sum games for finite horizon continuous-time Markov processes

This chapter considers a two-person zero-sum continuous-time Markov pure jump game in Borel state and action spaces over a fixed finite horizon. The main assumption on the model is the existence of a drift function, which bounds the reward rate. Under some regularity conditions, we show that the game has a value, and both of the players have their optimal policies. So there are two action spaces $\mathbf{A}$ for the maximizer and $\mathbf{B}$ for the minimizer. Also π denotes the policy for maximizer and ψ denotes the policy for minimizer. Other definitions are the same as previous replacing one action with two. Π and Ψ denote the classes of policies for the maximizer and minimizer respectively.

7.1 Model description

Now let $T \in (0, \infty)$ be a fixed time duration, and define

$$W(x, \pi, \psi) := E_x^{\pi,\psi}\left[\int_0^T \int_{\mathbf{A}\times\mathbf{B}} r(t, \xi_t, a, b)\pi(da|\omega, t)\psi(db|\omega, t)dt\right] + E_x^{\pi,\psi}\left[g(T, \xi_T)\right]$$

for each $(\pi, \psi) \in \Pi \times \Psi$, and $x \in \mathbf{S}$. The conditions to be imposed below assure that the above expectations are finite, see Lemma 7.1.

The lower value of the zero-sum continuous-time Markov pure jump game over the fixed horizon $[0, T]$ is defined by

$$L(x) := \sup_{\pi\in\Pi} \inf_{\psi\in\Psi} W(x, \pi, \psi), \ \forall\ x \in \mathbf{S},$$

and the upper value is defined by

$$U(x) := \inf_{\psi\in\Psi} \sup_{\pi\in\Pi} W(x, \pi, \psi), \ \forall\ x \in \mathbf{S}.$$

Apparently, $U(x) \geq L(x)$ for each $x \in \mathbf{S}$. If $U(x) = L(x)$ for each $x \in \mathbf{S}$, the function $\mathcal{W}$ defined by their common values is called the value of the game.

Definition 7.1. A policy $\pi^* \in \Pi$ is called optimal for the maximizer if it satisfies that $\inf_{\psi \in \Psi} W(x, \pi^*, \psi) = U(x)$ for each $x \in \mathbf{S}$. A policy $\psi^* \in \Psi$ is called optimal for the minimizer if $\sup_{\pi \in \Pi} W(x, \pi, \psi^*) = L(x)$ for each $x \in \mathbf{S}$.

It follows that the pair of optimal policies (π^*, ψ^*) in the above definition satisfies

$$U(x) = \inf_{\psi \in \Psi} W(x, \pi^*, \psi) \leq W(x, \pi^*, \psi^*) \leq \sup_{\pi \in \Pi} W(x, \pi, \psi^*) = L(x), \ \forall \ x \in \mathbf{S}.$$

Then $U(x) = L(x)$ for each $x \in \mathbf{S}$, i.e., the value of the game exists, if both players have their own optimal policies.

The main objective of this chapter is to show, under some conditions, that the function V exists, and both players have an optimal policy.

7.2 Conditions and relevant facts

In this section, we present the conditions imposed on the continuous-time Markov pure jump game model, and formulate their relevant consequences.

Condition 7.1. There exist $[1, \infty)$-valued measurable functions w_0 and w_1 on $\mathbf{S}$ and real constants $c_0 > 0$, $c_1 > 0$, $M_0 > 0$ and $M_1 > 0$ such that the following assertions hold.

(a) For each $(t, x, a, b) \in \mathbb{K}$, $\int_{\mathbf{S}} w_0(y) q(dy|t, x, a, b) \leq c_0 w_0(x)$.

(b) For each $x \in \mathbf{S}$, $\bar{q}_x \leq M_0 w_0(x)$.

(c) For each $(t, x, a, b) \in \mathbb{K}$, $|r(t, x, a, b)| \leq M_0 w_0(x)$, $|g(t, x)| \leq M_0 w_0(x)$.

(d) For each $(t, x, a, b) \in \mathbb{K}$, $\int_{\mathbf{S}} w_1(y) q(dy|t, x, a, b) \leq c_1 w_1(x)$.

(e) For each $x \in \mathbf{S}$, $w_0(x) \bar{q}_x \leq M_1 w_1(x)$.

Lemma 7.1. Suppose Condition 7.1 is satisfied. Let some pair of policies $(\pi, \psi) \in \Pi \times \Psi$ be arbitrarily fixed. Then the following assertions hold.

(a) $P_x^{\pi,\psi}(t_\infty = \infty) = 1$ for each $x \in \mathbf{S}$.

(b) $E_x^{\pi,\psi}[w_0(\xi_t)] \leq e^{c_0 t} w_0(x)$ for each $t \geq 0$ and $x \in \mathbf{S}$.

(c) $|W(x, \pi, \psi)| \leq (T+1) M_0 e^{c_0 T} w_0(x)$ for each $x \in \mathbf{S}$.

(d) For each $u \in C^{1,0}_{w_0,w_1}([0,T] \times \mathbf{S})$,

$$\begin{aligned} & E_x^{\pi,\psi}\left[\int_0^T \left(u'(t,\xi_t) + \int_{\mathbf{S}}\int_{\mathbf{A}}\int_B u(t,x) q(dx|t,\xi_t,a,b)\pi(da|\omega,t)\psi(db|\omega,t)\right) dt\right] \\ = \;& E_x^{\pi,\psi}[u(T,\xi_T)] - u(0,x). \end{aligned}$$

for each $x \in \mathbf{S}$.

Proof. See Lemmas 3.1, 3.2 and 3.3 in [53]. □

Throughout the rest of this paper, let m be an $[1,\infty)$-valued measurable function on $\mathbf{S}$ such that $\bar{q}_x \leq m(x)$ for each $x \in \mathbf{S}$. Such a function exists by the Novikov seperation theorem, see [72]. We introduce the following stochastic kernel on $\mathbf{S}$ from $(t,x,a,b) \in \mathbb{K}$ defined by

$$\tilde{p}(dy|t,x,a,b) := \delta_x(dy) + \frac{q(dy|t,x,a,b)}{m(x)}, \ \forall\ (t,x,a,b) \in \mathbb{K}.$$

Condition 7.2. For each $t \in [0,T]$ and $x \in \mathbf{S}$,

(a) $r(t,x,a,b)$ is continuous in $(a,b) \in \mathbf{A}(t,x) \times \mathbf{B}(t,x)$; and

(b) for each measurable function u on $\mathbf{S}$ such that $\sup_{x\in\mathbf{S}} \frac{|u(x)|}{w_0(x)} < \infty$, $\int_{\mathbf{S}} u(y)\tilde{p}(dy|t,x,a,b)$ is continuous in $(a,b) \in \mathbf{A}(t,x) \times \mathbf{B}(t,x)$.

Suppose that Condition 7.1 is satisfied. For each $t \in [0,\infty)$, $x \in \mathbf{S}$, $\lambda \in$

$\mathbb{P}(\mathbf{A}(t,x))$ and $\mu \in \mathbb{P}(\mathbf{B}(t,x))$, we introduce the notations

$$\begin{aligned} q(dy|t,x,\lambda,\mu) &:= \int_{\mathbf{A}(t,x)} \int_{\mathbf{B}(t,x)} q(dy|t,x,a,b)\lambda(da)\mu(db), \\ r(t,x,\lambda,\mu) &:= \int_{\mathbf{A}(t,x)} \int_{\mathbf{B}(t,x)} r(t,x,a,b)\lambda(da)\mu(db). \end{aligned}$$

(In particular, the integral in the second line of the above is finite under Condition 7.1.) Then $q(dy|t,x,\lambda,\mu)$ and $r(t,x,\lambda,\mu)$ are measurable on $\mathcal{K}$, where

$$\mathcal{K} := \{(t,x,\lambda,\mu) \in [0,\infty) \times \mathbf{S} \times \mathbb{P}(\mathbf{A}) \times \mathbb{P}(\mathbf{B}) : \lambda \in \mathbb{P}(\mathbf{A}(t,x)),\ \mu \in \mathbb{P}(\mathbf{B}(t,x))\}.$$

In greater details, since $(t,x) \mapsto \mathbf{A}(t,x)$ and $(t,x) \mapsto \mathbf{B}(t,x)$ are measurable and compact-valued multifunctions, as assumed earlier, by Theorem 3 of [60] and Proposition 7.22 of [9], so are the multifunctions $(t,x) \mapsto \mathbb{P}(\mathbf{A}(t,x))$ and $(t,x) \mapsto \mathbb{P}(\mathbf{B}(t,x))$. It follows from Theorem 3 of [59] that $\mathcal{K}$ is measurable in the Borel space $[0,\infty) \times \mathbf{S} \times \mathbb{P}(\mathbf{A}) \times \mathbb{P}(\mathbf{B})$. By Corollary 7.29.1 and Lemma 7.21 of [9] that $q(dy|t,x,\lambda,\mu)$ and $r(t,x,\lambda,\mu)$ are measurable on $\mathcal{K}$.

The next lemma, used repeatedly in the next section, is known. But we include its rather short proof for completeness. Recall that $\mathbf{A}(t,x)$ and $\mathbf{B}(t,x)$ are compact subsets of $\mathbf{A}$ and $\mathbf{B}$ as assumed in the beginning of the model description.

Lemma 7.2. Suppose that Conditions 7.1 and 7.2 are satisfied.

(a) Let $t \in [0,T]$ and $x \in \mathbf{S}$ be arbitrarily fixed. For each $u \in B_{w_0}([0,T] \times \mathbf{S})$, the functions $r(t,x,\lambda,\mu)$ and $\int_{\mathbf{S}} u(t,y)q(dy|t,x,\lambda,\mu)$ are continuous in $(\lambda,\mu) \in \mathbb{P}(\mathbf{A}(t,x)) \times \mathbb{P}(\mathbf{B}(t,x))$.

(b) If a function $h(t,x,\lambda,\mu)$ is real-valued and measurable on $\mathcal{K}$, and continuous in $(\lambda,\mu) \in \mathbb{P}(\mathbf{A}(t,x)) \times \mathbb{P}(\mathbf{B}(t,x))$ (for each fixed $(t,x) \in [0,T] \times \mathbf{S}$), then the function

$$(t,x,\lambda) \to \inf_{\mu \in \mathbb{P}(\mathbf{B}(t,x))} h(t,x,\lambda,\mu)$$

is measurable on $\{(t,x,\lambda) \in [0,T] \times \mathbf{S} \times \mathbb{P}(\mathbf{A}) : \lambda \in \mathbb{P}(\mathbf{A}(t,x))\}$ and continuous in $\lambda \in \mathbb{P}(\mathbf{A}(t,x))$ (for each fixed $(t,x) \in [0,T] \times \mathbf{S}$).

Proof. (a) For the fixed $t \in [0,T]$ and $x \in \mathbf{S}$, the functions $r(t,x,a,b)$ and $\int_{\mathbf{S}} u(t,y)q(dy|t,x,a,b)$ are bounded and continuous in $(a,b) \in \mathbf{A}(t,x) \times \mathbf{B}(t,x)$. The statement follows from Corollary 7.29.1 and Lemma 7.12 of [9], and the Tietze extension theorem.

(b) The first assertion follows from Theorem 2 of [59]. The second assertion is a consequence of the Berge theorem, see Theorem 17.31 in [1]. □

7.3 Main statement

In this section, we present and prove the main result of this paper; see Theorem 7.1 below.

Under Conditions 7.1 and 7.2, it follows from Lemmas 7.1 and 7.2 and the fundamental theorem of calculus that the following operator G maps $u \in B_{w_0}([0,T] \times \mathbf{S})$ to $C^{1,0}_{w_0,w_1}([0,T] \times \mathbf{S})$:

$$\begin{aligned} & G[u](t,x) \\ := \;& e^{-m(x)(T-t)}g(T,x) + \int_0^{T-t} e^{-m(x)s} \\ & \sup_{\lambda \in \mathbb{P}(\mathbf{A}(t+s,x))} \inf_{\mu \in \mathbb{P}(\mathbf{B}(t+s,x))} \left\{ r(t+s,x,\lambda,\mu) + m(x) \int_{\mathbf{S}} u(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu) \right\} ds \end{aligned}$$

for each $t \in [0,T]$ and $x \in \mathbf{S}$.

Proposition 7.1. Suppose that Conditions 7.1 and 7.2 are satisfied. There is a fixed point of the operator G in $C^{1,0}_{w_0,w_1}([0,T] \times \mathbf{S})$.

Proof. Let us define

$$u_0(t,x) := \frac{M_0}{c_0} \left\{ c_0 e^{c_0(T-t)} + e^{c_0(T-t)} - 1 \right\} w_0(x) \geq 0$$

for each $t \in [0,T]$ and $x \in \mathbf{S}$. Then u_0 belongs to $C^{1,0}_{w_0,w_1}([0,T] \times \mathbf{S})$. For each $n \geq 0$, we legitimately define $u_{n+1} := G[u_n]$. The rest of the proof goes in two

steps.

Step 1. Show that $\{u_n\}$ is a monotone nonincreasing sequence, and for each $n = 0, 1, \ldots,$

$$|u_n(t,x)| \le u_0(t,x) = \frac{M_0}{c_0}\left\{c_0 e^{c_0(T-t)} + e^{c_0(T-t)} - 1\right\} w_0(x).$$

for each $t \in [0,T]$ and $x \in \mathbf{S}$.

For each $t \in [0,T]$ and $x \in \mathbf{S}$,

$$\begin{aligned}
& u_1(t,x) = G[u_0](t,x) \\
\le\ & e^{-m(x)(T-t)} M_0 w_0(x) + \int_0^{T-t} e^{-m(x)s} \\
& \sup_{\lambda \in \mathbb{P}(\mathbf{A}(t+s,x))} \inf_{\mu \in \mathbb{P}(\mathbf{B}(t+s,x))} \left\{ M_0 w_0(x) + m(x) \int_{\mathbf{S}} u_0(t+s,y) \tilde{p}(dy|t+s,x,\lambda,\mu) \right\} ds \\
=\ & e^{-m(x)(T-t)} M_0 w_0(x) + M_0 w_0(x) \int_0^{T-t} e^{-m(x)s} ds \\
& + \frac{M_0}{c_0} \int_0^{T-t} e^{-m(x)s} m(x) \left\{ c_0 e^{c_0(T-t-s)} + e^{c_0(T-t-s)} - 1 \right\} w_0(x) ds \\
& + \frac{M_0}{c_0} \int_0^{T-t} e^{-m(x)s} \sup_{\lambda \in \mathbb{P}(\mathbf{A}(t+s,x))} \inf_{\mu \in \mathbb{P}(\mathbf{B}(t+s,x))} \left\{ (c_0 e^{c_0(T-t-s)} + e^{c_0(T-t-s)} - 1) \right. \\
& \left. \int_{\mathbf{S}} w_0(y) q(dy|t+s,x,\lambda,\mu) \right\} ds \\
\le\ & e^{-m(x)(T-t)} M_0 w_0(x) + M_0 w_0(x) \int_0^{T-t} e^{-m(x)s} ds \\
& + \frac{M_0}{c_0} \int_0^{T-t} e^{-m(x)s} m(x) \left\{ c_0 e^{c_0(T-t-s)} + e^{c_0(T-t-s)} - 1 \right\} w_0(x) ds \\
& + \frac{M_0}{c_0} \int_0^{T-t} e^{-m(x)s} (c_0 e^{c_0(T-t-s)} + e^{c_0(T-t-s)} - 1) c_0 w_0(x) ds,
\end{aligned}$$

where the first and the last inequalities are by Condition 7.1. For the third

summand on the right hand side of the last inequality, integration by parts gives

$$\begin{aligned} & \frac{M_0}{c_0}\int_0^{T-t} e^{-m(x)s}m(x)\left\{c_0e^{c_0(T-t-s)}+e^{c_0(T-t-s)}-1\right\}w_0(x)ds \\ = \; & -w_0(x)M_0e^{-m(x)(T-t)}+\frac{M_0}{c_0}\left\{c_0e^{c_0(T-t)}+e^{c_0(T-t)}-1\right\}w_0(x) \\ & -\frac{M_0}{c_0}w_0(x)\int_0^{T-t} e^{-m(x)s}\left\{c_0^2e^{c_0(T-t-s)}+c_0e^{c_0(T-t-s)}\right\}ds. \end{aligned}$$

This, together with the previous calculations, shows that

$$u_1(t,x)\le \frac{M_0}{c_0}\left\{c_0e^{c_0(T-t)}+e^{c_0(T-t)}-1\right\}w_0(x)=u_0(t,x),\ \forall\ t\in[0,T],\ x\in\mathbf{S}.$$

It follows from this and the monotonicity of the operator G that $\{u_n\}$ is a monotone nonincreasing sequence, and for each $n\ge 0$,

$$u_n(t,x)\le u_0(t,x)=\frac{M_0}{c_0}\left\{c_0e^{c_0(T-t)}+e^{c_0(T-t)}-1\right\}w_0(x)$$

for each $t\in[0,T]$ and $x\in\mathbf{S}$.

On the other hand, a similar calculation to the above gives

$$\begin{aligned} & u_1(t,x) \\ \ge \; & -e^{-m(x)(T-t)}M_0w_0(x)-\int_0^{T-t} e^{-m(x)s} \\ & \sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\inf_{\mu\in\mathbb{P}(\mathbf{B}(t+s,x))}\left\{-M_0w_0(x)-m(x)\int_{\mathbf{S}}u_0(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu)\right\}ds \\ \ge \; & -u_0(t,x) \end{aligned}$$

for each $t\in[0,T]$ and $x\in\mathbf{S}$. Hence, for each $n\ge 0$,

$$|u_n(t,x)|\le u_0(t,x)=\frac{M_0}{c_0}\left\{c_0e^{c_0(T-t)}+e^{c_0(T-t)}-1\right\}w_0(x).$$

for each $t\in[0,T]$ and $x\in\mathbf{S}$.

Step 2. Consider the function u^* defined by $u^*(t,x):=\lim_{n\to\infty}u_n(t,x)$ for each $t\in[0,T]$ and $x\in\mathbf{S}$. The limit exists due to the monotone convergence. We

show that u^* is a fixed point of the operator G in $C^{1,0}_{w_0,w_1}([0,T]\times\mathbf{S})$.

It follows from the definition of u^* and what was established in Step 1 that

$$|u^*(t,x)| \le \frac{M_0}{c_0}\left\{c_0 e^{c_0(T-t)} + e^{c_0(T-t)} - 1\right\} w_0(x)$$

for each $t\in[0,T]$ and $x\in\mathbf{S}$, that is, $u^*\in B_{w_0}([0,T]\times\mathbf{S})$.

We verify that u^* is a fixed point of G as follows. It is evident that for each $n\ge 0$, $G[u^*](t,x)\le G[u_n](t,x) = u_{n+1}(t,x)$ for each $t\in[0,T]$ and $x\in\mathbf{S}$. Hence,

$$G[u^*](t,x) \le u^*(t,x) \tag{7.1}$$

for each $t\in[0,T]$ and $x\in\mathbf{S}$.

The rest of this proof mainly verifies the opposite direction of the above inequality. Let $x\in S$ be fixed, and consider the space of $\mathbb{P}(A)$-valued measurable mappings say λ on $[0,T]$ such that for each $t\in[0,T]$, $\lambda_t\in\mathbb{P}(A(t,x))$. We denote this space by $\mathcal{R}_{\mathbf{A}}$ and $\mathcal{R}_{\mathbf{B}}$ for the maximizer and minimizer respectively.

Note that by Theorem 2 of [59], applicable due to Lemma 7.2, for each $x\in\mathbf{S}$ and $t\in[0,T]$,

$$\begin{aligned}
&\int_0^{T-t} e^{-m(x)s} \\
&\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\inf_{\mu\in\mathbb{P}(\mathbf{B}(t+s,x))}\left\{r(t+s,x,\lambda,\mu) + m(x)\int_{\mathbf{S}} u(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu)\right\} ds \\
= &\int_0^{T-t} e^{-m(x)s} \\
&\sup_{\lambda_{t+s}\in\mathcal{R}_{\mathbf{A}}}\inf_{\mu_{t+s}\in\mathcal{R}_{\mathbf{B}}}\left\{r(t+s,x,\lambda_{t+s},\mu_{t+s}) + m(x)\int_{\mathbf{S}} u(t+s,y)\tilde{p}(dy|t+s,x,\lambda_{t+s},\mu_{t+s})\right\} ds.
\end{aligned}$$

Fix $(t,x)\in[0,T]\times\mathbf{S}$ and some $\mu_{t+s}\in\mathcal{R}_{\mathbf{B}}$ arbitrarily. By Theorem 2 of [59] and Lemma 7.2, for each $n\ge 0$, there exists $\lambda^n_{t+s}\in\mathcal{R}_{\mathbf{A}}$ (λ^n depends also on

(t,x))such that

$$\begin{aligned} & u_{n+1}(t,x) = G[u_n](t,x) \\ \leq\ & e^{-m(x)(T-t)}g(T,x) + \int_0^{T-t} e^{-m(x)s} \\ & \left\{ r(t+s,x,\lambda^n_{t+s},\mu_{t+s}) + m(x)\int_{\mathbf{S}} u_n(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s}) \right\} ds \end{aligned} \tag{7.2}$$

Recall that $\mathbf{A}(t,x) \subseteq \mathbf{A}(x)$ for each $x \in \mathbf{S}$ and $t \in [0,\infty)$. Since $\mathcal{R}_{\mathbf{A}}$ is compact metrizable, without loss of generality we assume that the sequence $\{\lambda^n\}$ in $\mathcal{R}_{\mathbf{A}}$ converges to some $\lambda^* \in \mathcal{R}_{\mathbf{A}}$, for otherwise one can take a convergent subsequence and relabel it. Note that

$$\begin{aligned} & \left| \int_0^{T-t} e^{-m(x)s} m(x) \int_{\mathbf{S}} u_n(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds \right. \\ & \left. - \int_0^{T-t} e^{-m(x)s} m(x) \int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds \right| \\ \leq\ & \int_0^{T-t} e^{-m(x)s} m(x) \int_{\mathbf{A}(t+s,x)} \int_{\mathbf{S}} |u_n(t+s,y) - u^*(t+s,y)| \tilde{p}(dy|t+s,x,a,\mu_{t+s})\lambda^n_{t+s}(da)ds \\ \leq\ & \int_0^{T-t} e^{-m(x)s} m(x) \sup_{a\in\mathbf{A}(t+s,x)} \left\{ \int_{\mathbf{S}} |u_n(t+s,y) - u^*(t+s,y)| \tilde{p}(dy|t+s,x,a,\mu_{t+s}) \right\} ds. \end{aligned} \tag{7.3}$$

On the other hand,

$$\begin{aligned} & \lim_{n\to\infty} \sup_{a\in\mathbf{A}(t+s,x)} \left\{ \int_{\mathbf{S}} |u_n(t+s,y) - u^*(t+s,y)| \tilde{p}(dy|t+s,x,a,\mu_{t+s}) \right\} \\ =\ & \sup_{a\in\mathbf{A}(t+s,x)} \left\{ \lim_{n\to\infty} \int_{\mathbf{S}} |u_n(t+s,y) - u^*(t+s,y)| \tilde{p}(dy|t+s,x,a,\mu_{t+s}) \right\} \\ =\ & 0, \end{aligned}$$

where the first equality is by Theorem A.1.5 of [6], applicable under Condition 7.2, and the last equality is by the dominated convergence theorem, applicable under Condition 7.1. It follows from this, (7.3) and the dominated convergence

theorem that

$$\lim_{n\to\infty}\left|\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u_n(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds\right.$$
$$\left.-\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds\right| = 0.$$

Now as $n\to\infty$,

$$\begin{aligned}
&\left|\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u_n(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds\right.\\
&\left.-\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda^*_{t+s},\mu_{t+s})ds\right|\\
\leq\ &\left|\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u_n(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds\right.\\
&\left.-\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds\right|\\
&+\left|\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda^n_{t+s},\mu_{t+s})ds\right.\\
&\left.-\int_0^{T-t} e^{-m(x)s}m(x)\int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda^*_{t+s},\mu_{t+s})ds\right|\\
\to\ &0,
\end{aligned}$$

where the convergence to zero is also by the definition of the Young topology. It follows from this and the definition of the Young topology again that, after passing to the limit as $n\to\infty$ on the both sides of (7.2),

$$\begin{aligned}
&u^*(t,x)\\
\leq\ &e^{-m(x)(T-t)}g(T,x)+\int_0^{T-t} e^{-m(x)s}\left\{r(t+s,x,\lambda^*_{t+s},\mu_{t+s})\right.\\
&\left.+m(x)\int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda^*_{t+s},\mu_{t+s})\right\}ds\\
\leq\ &e^{-m(x)(T-t)}g(T,x)+\int_0^{T-t} e^{-m(x)s}\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\{r(t+s,x,\lambda,\mu_{t+s})\\
&\left.+m(x)\int_{\mathbf{S}} u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu_{t+s})\right\}ds. \qquad (7.4)
\end{aligned}$$

By Theorem 2 of [59], applicable due to Lemma 7.2, there exists $\mu^* \in \mathcal{R}_{\mathbf{B}}$ such that

$$\begin{aligned}
&\inf_{\mu\in\mathbb{P}(\mathbf{B}(t+s,x))}\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\left\{r(t+s,x,\lambda,\mu)+m(x)\int_{\mathbf{S}}u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu)\right\}\\
=\;&\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\left\{r(t+s,x,\lambda,\mu^*_{t+s})+m(x)\int_{\mathbf{S}}u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu^*_{t+s})\right\}
\end{aligned}$$

for each $s \in [0, T-t]$. By the Ky Fan minimax theorem, see Theorem 2 of [31],

$$\begin{aligned}
&\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\inf_{\mu\in\mathbb{P}(\mathbf{B}(t+s,x))}\left\{r(t+s,x,\lambda,\mu)+m(x)\int_{\mathbf{S}}u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu)\right\}\\
=\;&\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\left\{r(t+s,x,\lambda,\mu^*_{t+s})+m(x)\int_{\mathbf{S}}u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu^*_{t+s})\right\}
\end{aligned}$$

for each $s \in [0, T-t]$. Since $\mu \in \mathcal{R}_{\mathbf{B}}$ in (7.4) was arbitrarily fixed, we see from (7.4) and the previous equality that

$$\begin{aligned}
&u^*(t,x)\\
\le\;& e^{-m(x)(T-t)}g(T,x)+\int_0^{T-t}e^{-m(x)s}\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\big\{r(t+s,x,\lambda,\mu^*_{t+s})\\
&+m(x)\int_{\mathbf{S}}u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu^*_{t+s})\Big\}\,ds\\
=\;& e^{-m(x)(T-t)}g(T,x)+\int_0^{T-t}e^{-m(x)s}\\
&\sup_{\lambda\in\mathbb{P}(\mathbf{A}(t+s,x))}\inf_{\mu\in\mathbb{P}(\mathbf{B}(t+s,x))}\left\{r(t+s,x,\lambda,\mu)+m(x)\int_{\mathbf{S}}u^*(t+s,y)\tilde{p}(dy|t+s,x,\lambda,\mu)\right\}ds\\
=\;& G[u^*](t,x).
\end{aligned}$$

Since $(t,x) \in [0,T]\times\mathbf{S}$ was arbitrarily fixed, this and (7.1) imply

$$u^*(t,x) = G[u^*](x,t),\ \forall\ t\in[0,T],\ x\in\mathbf{S}.$$

Finally, since $u^* \in B_{w_0}([0,T]\times\mathbf{S})$, and G maps each element of $B_{w_0}([0,T]\times\mathbf{S})$ to $C^{1,0}_{w_0,w_1}([0,T]\times\mathbf{S})$ as mentioned earlier, it follows that u^* is a fixed point of G in $C^{1,0}_{w_0,w_1}([0,T]\times\mathbf{S})$. □

Theorem 7.1. Suppose that Conditions 7.1 and 7.2 are satisfied. Then the zero-sum continuous-time Markov pure jump game has a value V, and both the

maximizer and minimizer have an optimal Markov policy. In particular, there is a pair of Markov policies $(\pi_*^M, \psi_*^M) \in \Pi \times \Psi$ such that $W(x, \pi_*^M, \psi_*^M) = V(x)$ for each $x \in \mathbf{S}$.

Proof. By Proposition 7.1, we can consider a solution $u \in C^{1,0}_{w_0,w_1}([0,T] \times \mathbf{S})$ to the following equation

$$\begin{aligned} & u(t,x) \\ = \; & e^{-m(x)(T-t)} g(T,x) + \int_0^{T-t} e^{-m(x)s} \\ & \sup_{\lambda \in \mathbb{P}(\mathbf{A}(t+s,x))} \inf_{\mu \in \mathbb{P}(\mathbf{B}(t+s,x))} \left\{ r(t+s,x,\lambda,\mu) + m(x) \int_{\mathbf{S}} u(t+s,y) \tilde{p}(dy|t+s,x,\lambda,\mu) \right\} ds, \\ & \forall\ t \in [0,T],\ x \in \mathbf{S}. \end{aligned}$$

Then

$$\begin{aligned} & e^{-m(x)t} u(t,x) \\ = \; & e^{-m(x)T} g(T,x) + \int_0^{T-t} e^{-m(x)(t+s)} \\ & \sup_{\lambda \in \mathbb{P}(\mathbf{A}(t+s,x))} \inf_{\mu \in \mathbb{P}(\mathbf{B}(t+s,x))} \left\{ r(t+s,x,\lambda,\mu) + m(x) \int_{\mathbf{S}} u(t+s,y) \tilde{p}(dy|t+s,x,\lambda,\mu) \right\} ds \\ = \; & e^{-m(x)T} g(T,x) + \int_t^{T} e^{-m(x)(s)} \\ & \sup_{\lambda \in \mathbb{P}(\mathbf{A}(s,x))} \inf_{\mu \in \mathbb{P}(\mathbf{B}(s,x))} \left\{ r(s,x,\lambda,\mu) + m(x) \int_{\mathbf{S}} u(s,y) \tilde{p}(dy|s,x,\lambda,\mu) \right\} ds \\ & \forall\ t \in [0,T],\ x \in \mathbf{S}. \end{aligned}$$

It follows that for each $x \in \mathbf{S}$,

$$u(T,x) = g(T,x) \tag{7.5}$$

and

$$u'(t,x) + \sup_{\lambda \in \mathbb{P}(\mathbf{A}(t,x))} \inf_{\mu \in \mathbb{P}(\mathbf{B}(t,x))} \left\{ r(t,x,\lambda,\mu) + \int_{\mathbf{S}} u(s,y) q(dy|t,x,\lambda,\mu) \right\} = 0$$

almost everywhere on $[0,T]$.

By Theorem 2 of [59], applicable due to Lemma 7.2, there exists a Markov

policy say π_*^M for the maximizer such that for each $x \in \mathbf{S}$,

$$\begin{aligned} & u'(t,x) + \inf_{\mu \in \mathbb{P}(\mathbf{B}(t,x))} \left\{ \int_{\mathbf{A}} r(t,x,a,\mu) \pi_*^M(da|x,t) + \int_{\mathbf{S}} u(t,y) \int_{\mathbf{A}} q(dy|t,x,a,\mu) \pi_*^M(da|x,t) \right\} \\ = \; & 0 \end{aligned}$$

almost everywhere on $[0,T]$, that is, for each $\mu \in \mathbb{P}(\mathbf{B}(t,x))$,

$$u'(t,x) + \int_{\mathbf{A}} r(t,x,a,\mu) \pi_*^M(da|x,t) + \int_{\mathbf{S}} u(t,y) \int_{\mathbf{A}} q(dy|t,x,a,\mu) \pi_*^M(da|x,t) \geq 0$$

almost everywhere on $[0,T]$.

Now, by Lemma 7.1(d), for each policy $\psi \in \Psi$ for the minimizer and $x \in \mathbf{S}$,

$$\begin{aligned} & E_x^{\pi_*^M, \psi}[g(T, \xi_T)] - u(0,x) = E_x^{\pi_*^M, \psi}[u(T, \xi_T)] - u(0,x) \\ = \; & E_x^{\pi_*^M, \psi} \left[\int_0^T \left(u'(t,\xi_t) + \int_{\mathbf{S}} \int_{\mathbf{A}} \int_{\mathbf{B}} u(t,x) q(dx|t,\xi_t,a,b) \pi_*^M(da|\xi_t,t) \psi(db|\omega,t) \right) dt \right] \\ \geq \; & -E_x^{\pi_*^M, \psi} \left[\int_0^T \int_{\mathbf{A}} \int_{\mathbf{B}} r(t,\xi_t,a,b) \pi_*^M(da|\xi_t,t) \psi(db|\omega,t) dt \right], \end{aligned}$$

where the first equality is by (7.5). That is,

$$u(0,x) \leq W(x, \pi_*^M, \psi), \; \forall \; x \in \mathbf{S}.$$

Since $\psi \in \Psi$ was arbitrarily fixed, we see

$$u(0,x) \leq \inf_{\psi \in \Psi} W(x, \pi_*^M, \psi) \leq \sup_{\pi \in \Pi} \inf_{\psi \in \Psi} W(x, \pi, \psi) = L(x), \; \forall \; x \in \mathbf{S}. \tag{7.6}$$

Similarly, by By Theorem 2 of [59] and the Ky Fan minimax theorem (see Theorem 2 of [31]), there exists a Markov policy say ψ_*^M for the minimizer such that for each $x \in \mathbf{S}$,

$$\begin{aligned} & u'(t,x) + \sup_{\lambda \in \mathbb{P}(\mathbf{A}(t,x))} \left\{ \int_{\mathbf{B}} r(t,x,\lambda,b) \psi_*^M(db|x,t) + \int_{\mathbf{S}} u(t,y) \int_{\mathbf{B}} q(dy|t,x,\lambda,b) \psi_*^M(db|x,t) \right\} \\ = \; & 0 \end{aligned}$$

almost everywhere on $[0,T]$. Then by using Lemma 7.1(d), one can show as in

the above that

$$u(0,x) \geq \sup_{\pi\in\Pi} W(x,\pi,\psi_*^M) \geq \inf_{\psi\in\Psi}\sup_{\pi\in\Pi} W(x,\pi,\psi) = U(x),\ \forall\ x \in \mathbf{S}.$$

Combining this and (7.6) yields

$$u(0,x) = L(x) = U(x) = \sup_{\pi\in\Pi} W(x,\pi,\psi_*^M) = \inf_{\psi\in\Psi} W(x,\pi_*^M,\psi) = W(x,\pi_*^M,\psi_*^M).$$

The proof is completed. □

References

[1] Aliprantis, C. and Border, K. (2006). Infinite Dimensional Analysis. *Springer, New York.*

[2] Altman, E. (1999). Constrained Markov Decision Processes. *Chapman and Hall/CRC, Boca Raton.*

[3] Anderson, W. J. (1991). Continuous-Time Markov Chains. *New York: Springer-Verlag.*

[4] Bäuerle, N. and Jaśkiewicz, A. (2015). Risk-sensitive divident problems. *Eur. J. Oper. Res.* **242**, 161–171.

[5] Bäuerle, N. and Rieder, U. (2009). MDP algorithms for portfolio optimization problems in pure jump markets. *Finance Stoch.* **13**, 591–611.

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

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

[8] Bäuerle, N. and Popp, A. (2018). Risk-sensitive stopping problems for continuous-time Markov chains. *Stochastics* **90**, 411–431.

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

[10] Blok, H. and Spieksma, F. (2015). Countable state Markov decision processes with unbounded jump rates and discounted optimality equation and approximations. *Adv. Appl. Probab.* **47**, 1088–1107.

[11] Cavazos-Cadena, R. and Hernndez-Hernndez, D. (2011). Discounted approximations for risk-sensitive average criteria in Markov decision chains with finite state space. *Math. Oper. Res.* **36**, 133–146.

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

[13] Cavazos-Cadena, R. and Montes-de-Oca, R. (2000). Nearly optimal policies in risk-sensitive positive dynamic programming on discrete spaces. *Math. Meth. Oper. Res.* **52**, 133–167.

[14] Chen, A., Pollett, P., Zhang, H. and Cairns, B. (2005). Uniqueness criteria for continuous-time Markov chains with general transition structures. *Adv. Appl. Probab.* **37**, 1056–1074.

[15] Chen, M. (2015). Practical criterion for uniqueness of q-processes. *Chinese J. Appl. Probab. Stat.* **31**, 213–224.

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

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

[18] Costa, O. and Davis, M. (1989). Impulsive control of piecewise-deterministic processes. *Math. Control Signals Systems* **2**, 187–206.

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

[20] Costa, O. and Dufour, F. (2015). A linear programming formulation for constrained discounted continuous control for piecewise deterministic Markov processes. *J. Math. Anal. Appl.* **424**, 892–914.

[21] Costa, O. and Raymundo, C. (2000). Impulse and continuous control of piecewise deterministic Markov processes. *Stochastics* **70**, 75–107.

[22] Couwenbergh, H. (1980). Stochastic games with metric state space. *Int. J. Game Theory* **9**, 25–36.

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

[24] Di Masi, G. and Stettner, L. (1999). Risk-sensitive control of discrete-time Markov processes with infinite horizon. *SIAM J. Control Optim.* **38**, 61–78.

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

[26] Dufour, F., Horiguchi, M. and Piunovskiy, A. (2012). The expected total cost criterion for Markov decision processes under constraints: a convex analytic approach. *Adv. Appl. Probab.* **44**, 774–793.

[27] Dufour, F. and Piunovskiy, A. (2013). The expected total cost criterion for Markov decision processes under constraints. *Adv. Appl. Probab.* **45**, 837–859.

[28] Dufour, F. and Piunovskiy, A. (2015). Impulsive control for continuous-time Markov decision processes. *Adv. Appl. Probab.* **47**, 106–127.

[29] Dufour, F. and Prieto-Rumean, T. (2016). Conditions for the solvability of the linear programming formulation for constrained discounted Markov decision processes. *Appl. Math. Optim.* **74**, 27–51.

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

[31] Fan, K. (1953). Minimax theorems. *Proc. Natl. Acad. Sci. USA*. **39**, 42–47.

[32] Feinberg, E. (1982). Controlled Markov processes with arbitrary numerical criteria. *Theory Probab. Appl.* **27**, 486–503.

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

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

[35] Feinberg, E. (2012). Reduction of discounted continuous-time MDPs with unbounded jump and reward rates to discrete-time total-reward MDPs. *Optimization, Control, and Applications of Stochastic Systems*, Hernandez-Hernandez, D. and Minjarez-Sosa, A. (eds): 77–97, Birkhäuser, Bassel.

[36] Feinberg, E., Mandava, M. and Shiryaev, A. (2017) Kolmogorov's equations for jump Markov processes with unbounded jump rates. *Ann. Oper. Res.*, 1–18.

[37] Feinberg, E., Mandava, M. and Shiryaev, A. (2013). Sufficiency of Markov policies for continuous-time Markov decision processes and solutions of Kolmogorov's forward equation for jump Markov processes. *Proc. 52nd IEEE CDC*, 5728-5732. Dec, 2013, Florence, Italy.

[38] Feinberg, E., Mandava, M. and Shiryaev, A. (2014). On solutions of Kolmogorov's equations for nonhomogeneous jump Markov processes *J. Math. Anal. Appl.* **411**, 261–270,

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

[40] Feinberg, E. and Sonin, I. (1996). Notes on equivalent stationary policies in Markov decision processes with total rewards. *Math. Meth. Oper. Res.* **44**, 205–221.

[41] Filar, J. and Vrieze, K. (1997). Competitive Markov Decision Processes. *Springer, New York.*

[42] 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.

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

[44] Gihman, I. and Skorohod, A. (1975). The Theory of Stochastic Processes II. *Springer, Berlin.*

[45] Guo, X., Piunovskiy, A. and Zhang, Y. (2017). Note on discounted continuous-time Markov decision processes with a lower bounding function. *J. Appl. Prob.* **54**, 1071–1088.

[46] Guo, X. and Zhang, Y. (2018) On risk-sensitive piecewise deterministic Markov decision processes. *Appl. Math. Optim.*, in press, https://doi.org/10.1007/s00245-018-9485-x.

[47] Guo, X., Liu, Q.L. and Zhang, Y. (2018). Finite horizon risk-sensitive continuous-time Markov decision processes with unbounded transition and cost rates, *4OR-Q. J. Oper. Res.* **17**, 427–442.

[48] Guo, X.P. (2007). Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces. *Math. Oper. Res.*, **32**, 73–87.

[49] Guo, X.P. and Hernández-Lerma, O. (2007). Zero-sum games for continuous-time jump Markov processes in Polish spaces: discounted payoffs. *Adv. Appl. Probab.* **39**, 645–668.

[50] Guo, X.P. and Hernández-Lerma, O. (2009). Continuous-Time Markov Decision Processes. *Berlin: Springer-Verlag.*

[51] Guo, X.P. and Hernández-Lerma, O. (2011). New optimality conditions for average-payoff continuous-time Markov games in Polish spaces. *Sci. China Math.* **54**, 793–816.

[52] Guo, X.P., Huang, X.X. and Huang, Y.H. (2015). Finite-horizon optimality for continuous-time Markov decision processes with unbounded transition rates. *Adv. in Appl. Probab.* **47**, 1064–1087.

[53] Guo, X.P, Huang, Y.H. and Zhang, Y. (2016). Constrained continuous-time Markov decision processes on the finite horizon. *Appl. Math. Optim.*, **75**, 317–341.

[54] Guo, X.P. and Piunovskiy, A. (2011). Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. *Math. Oper. Res.* **36**, 105–132.

[55] Guo, X.P. and Song, X.Y. (2011). Discounted continuous-time constrained Markov decision processes in Polish spaces. *Ann. Appl. Probab.* **21**, 2016–2049.

[56] Guo, X.P. and Zhang, Y. (2017). Constrained total undiscounted continuous-time Markov decision processes. *Bernoulli* **23**, 1694–1736.

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

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

[59] Himmelberg, C., Parthasarathy, T., and Van Vleck, F. (1976). Optimal plans for dynamic programming problems. *Math. Oper. Res.* **1**, 390–394.

[60] Himmelberg, C. and Van Vleck, F. (1975). Multifunctions with values in a space of probability measures. *J. Math. Anal. Appl.* **50**, 108–112.

[61] Hordijk, A. and Van der Duyn Shouten, F. (1984). Discretization and weak convergence in Markov decision drift processes. *Math. Oper. Res.* **9**, 121–141.

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

[63] Huang, Y.H., Guo, X.P. and Lian, Z.T. (2019). Risk-sensitive finite-horizon piecewise deterministic Markov decision processes, *Oper. Res. Lett.*, **48**, 96–103.

[64] Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon-Nicodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 235–253.

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

[66] Jaskiewicz, A. (2007). Average optimality for risk-sensitive control with general state space. *Ann. Appl. Probab.* **17**, 654–675.

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

[68] Jaśkiewicz, A. and Nowak, A. (2011). Stochastic games with unbounded payoffs: applications to robust control in economics. *Dyn. Games Appl.* **1**, 253–279.

[69] Jaśkiewicz, A. and Nowak, A. (2011). Discounted dynamic programming with unbouned returns: application to economic models. *J. Math. Anal. Appl.* **378**, 450–462.

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

[71] Kitaev, M. and Rykov, V. (1995). Controlled Queueing Systems. *CRC Press, New York.*

[72] Kechris, A. (1995). Classical Descriptive Set Theory. *Springer, New York.*

[73] Kumar, K.S. and Chandan, P. (2013). Risk-sensitive control of jump process on denumerable state space with near monotone cost. *Appl. Math. Optim.* **68**, 311–331.

[74] Kumar, K.S. and Chandan, P. (2015). Risk-sensitive ergodic control of continuous-time Markov processes with denumerable state space. *Stochastic Anal. Appl.* **33**, 863–881.

[75] Kumar, P. and Shiau, T. (1981). Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games. *SIAM J. Control Optim.* **19**, 617–634.

[76] Kuznetsov, S. (1984). Inhomogeneous Markov processes. *J. Soviet Math.* **25**, 1380–1498.

[77] Maitra, A. and Parthasarathy, T. (1970). On stochastic games. *J. Optim. Theory Appl.* **5**, 289–300.

[78] Maitra, A. and Parthasarathy, T. (1971). On stochastic games, II. *J. Optim. Theory Appl.* **8**, 154–160.

[79] Miller, A., Miller, B. and Stepanyan, K. (2018). Joint continuous and impulsive control of Markov chains. In Proceedings of *26th Mediterranean Conference on Control and Automation*, Zadar, Croatia.

[80] Miller, B. (1968). Finite state continuous time Markov decision processes with an infinite planning horizon. *J. Math. Anal. Appl.* **22**, 552–569.

[81] Neyman, A. and Sorin, S. (2003) Stochastic Games and Applications. *Kluwer Academic Publishers, Dordrecht.*

[82] Nowak, A. (1985). Universally measurable strategies in zero-sum stochastic games. *Ann. Probab.* **13**, 269–287.

[83] Palczewski, J. and Stettner, L. (2017). Impulse control maximising average cost per unit time: a non-uniformly ergodic case. *SIAM J. Control Optim.* **55**, 936–960.

[84] Parthasarathy, T. (1973). Discounted, positive and noncooperative stochastic games. *Int. J. Game Theory* **2**, 25–37.

[85] Patek, S.(2001). On terminating Markov decision processes with a risk-averse objective function. *Automatica* **37**, 1379-1386.

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

[87] Piunovskiy, A. (2015). Randomized and relaxed strategies in continuous-time Markov decision processes. *SIAM J. Control Optim.* **53**, 3503–3533.

[88] Piunovskiy, 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.

[89] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim.* **49**, 2032–2061.

[90] Piunovskiy, A. and Zhang, Y. (2014). Discounted continuous-time Markov decision processes with unbounded rates and randomized history-dependent policies: the dynamic programming approach. *4OR-Q J. Oper. Res.* **12**, 49–75.

[91] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). Selected Topics in Continuous-Time Controlled Markov Chains and Markov Games. *Imperial College Press, London.*

[92] Rykov, V. (1966). Markov decision processes with finite state and decision spaces. *Theory Probab. Appl.* **11**, 302–311.

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

[94] Shapley, L. (1953). Stochastic games. *Proc. Natl. Acad. Sci. USA* **39**, 1095–1100.

[95] Spieksma, F. (2016). Kolmogorov forward equation and explosiveness in countable state Markov processes. *Ann. Oper. Res.* **241**, 3–22.

[96] Spieksma, F. (2015). Countable state Markov processes: non-explosiveness and moment function. *Probab. Eng. Inform. Sc.* **29**, 623–637.

[97] Van, C. and Morhaim, L. (2002). Optimal growth models with bounded or unbounded returns: a unifying approach. *J. Econom. Theory* **105**, 158–187.

[98] Van der Duyn Schouten, F. (1983). Markov Decision Processes with Continuous Time Parameter. *Mathematisch Centrum, Amsterdam.*

[99] Van der Wal, J. (1980). Stochastic Dynamic Programming: Successive Approximations and Nearly Optimal Strategies for Markov Decision Processes and Markov Games. *Mathematisch Centrum, Amsterdam.*

[100] Veinott, A. (1969). Discrete dynamic programming with sensitive discount optimality criteria. *Ann. Math. Stat.* **40**, 1635–1660.

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

[102] Wei, Q. and Chen, X. (2016). Continuous-time Markov decision processes under the risk-sensitive average cost criterion. *Oper. Res. Lett.* **44**, 457–462.

[103] Wei, Q. and Chen, X. (2016) Stochastic games for continuous-time jump processes under finite-horizon payoff criterion. *Appl. Math. Optim.* **74**, 273–301.

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

[105] Yushkevich, A. (1988). Bellman inequalities in Markov decision dterministic drift processes. *Stochastics* **23**, 25–77.

[106] Zachrisson L. (1964). Markov games. *Advances in Game Theory*, 210–253. Dresher M., L. Shapley and A. Tucker (eds), Princeton University Press, Princeton.

[107] Zhang, Y. (2017). On the nonexplosion and explosion for nonhomogeneous Markov pure jump processes. *J. Theor. Probab.* **31**, 1322–1355.

[108] Zhang, Y. (2017). Continuous-time Markov decision processes with exponential utility. *SIAM J. Control Optim.* **55**, 2636–2660.

A Q-function

The notations used in the Appendix are independent of the previous chapters. Now we give some facts for ease of reading. A (Borel-measurable) signed kernel $q(dy|x,s)$ on $\mathcal{B}(\mathbf{S})$ from $\mathbf{S}\times[0,\infty)$ is called a (conservative stable) Q-function on the Borel space $\mathbf{S}$ if the following conditions are satisfied.

(a) For each $s\geq 0$, $x\in\mathbf{S}$ and $\Gamma\in\mathcal{B}(\mathbf{S})$ with $x\notin\Gamma$, $\infty > q(\Gamma|x,s)\geq 0$.

(b) For each $(x,s)\in\mathbf{S}\times[0,\infty)$, $q(\mathbf{S}|x,s)=0$.

(c) For each $x\in\mathbf{S}$, $\sup_{s\in[0,\infty)}\{q(\mathbf{S}\setminus\{x\}|x,s)\}<\infty$.

For each Q-function q on $\mathbf{S}$, we put $\widetilde{q}(\Gamma|x,s) := q(\Gamma\setminus\{x\}|x,s)$, and $q_x(s) := \widetilde{q}(\mathbf{S}|x,s)$.

Given a Q-function q on S from $S\times[0,\infty)$, for each $\Gamma\in\mathcal{B}(S)$, $x\in S$, $s,t\in[0,\infty)$ and $s\leq t$, one can define

$$p_q^{(0)}(s,x,t,\Gamma) := \delta_x(\Gamma)e^{-\int_s^t q_x(v)dv},$$
$$p_q^{(n+1)}(s,x,t,\Gamma) := \int_s^t e^{-\int_s^u q_x(v)dv}\left(\int_S p_q^{(n)}(u,z,t,\Gamma)\widetilde{q}(dz|x,u)\right)du,\ \forall\ n=0,1,\ldots.$$

It is clear that one can legitimately define the sub-stochastic kernel $p_q(s,x,t,dy)$ on S by

$$p_q(s,x,t,\Gamma) := \sum_{n=0}^{\infty} p_q^{(n)}(s,x,t,\Gamma) \tag{A.1}$$

for each $x\in S$, $s,t\in[0,\infty)$, $s\leq t$, and $\Gamma\in\mathcal{B}(S)$. This is the Feller's construction for a transition function, i.e., p_q satisfies

$$p_q(s,x,s,dy)=\delta_x(dy)$$

and the Kolmogorov-Chapman equation

$$\int_S p_q(s,x,t,dy)p_q(t,y,u,\Gamma) = p_q(s,x,u,\Gamma),\ \forall\ \Gamma \in \mathcal{B}(S)$$

is valid for each $0 \le s \le t \le u < \infty$.

Condition A.1. *There exist a monotone nondecreasing sequence* $\{\widetilde{V}_n\}_{n=1}^{\infty} \subseteq [0,\infty) \times S$ *and a* $[0,\infty)$*-valued measurable function* $\widetilde{w}$ *on* $[0,\infty) \times S$ *such that the following hold.*

(a) As $n \uparrow \infty$, $\widetilde{V}_n \uparrow [0,\infty) \times S$.

(b) For each $n = 1, 2, \ldots$, $\sup_{x \in \hat{V}_n,\ t \in [0,\infty)} q_x(t) < \infty$, *where* $\hat{V}_n$ *denotes the projection of* $\widetilde{V}_n$ *on* S.

(c) As $n \uparrow \infty$, $\inf_{(t,x) \in ([0,\infty) \times S) \setminus \widetilde{V}_n} \widetilde{w}(t,x) \uparrow \infty$.

(d) For some constant $\rho' \in (0,\infty)$, *for each* $x \in S$ *and* $v \in [0,\infty)$,

$$\int_0^{\infty} \int_S \widetilde{w}(t+v,y) e^{-\rho' t - \int_0^t q_x(s+v)ds} \widetilde{q}(dy|x,t+v)dt \le \widetilde{w}(v,x).$$

The next statement follows from Theorem 3.2 of [107].

Theorem A.1. *If Condition A.1 is satisfied, then* $p_q(s,x,t,S) = 1$ *for each* $x \in S$, $s,t \in [0,\infty)$ *such that* $s \le t$.

B Risk-sensitive DTMDP

For ease of reference, we present the relevant notations and facts about the risk-sensitive problem for a DTMDP. The proofs of the presented statements can be found in [67] or [108]. Standard description of a DTMDP can be found in e.g., [57, 86].

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

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

Let us denote for each $n = 1, 2, \ldots, \infty$, $\mathbf{H}_n := \mathbf{S} \times (\mathbf{A} \times \mathbf{S})^n$ and $\mathbf{H}_0 := \mathbf{S}$. 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 $n = 0, 1, 2, \ldots$. A strategy $\sigma = (\sigma_n)$ is called deterministic Markov if for each $n = 0, 1, 2, \ldots$, $\sigma_n(da|h_n) = \delta_{\{\varphi_n(x_n)\}}(da)$, where φ_n is an $\mathbf{A}$-valued measurable mapping on $\mathbf{S}$. We identify such a deterministic Markov strategy with (φ_n). A deterministic Markov strategy (φ_n) is called deterministic stationary if φ_n does not depend on n, and it is identified with the underlying measurable mapping φ from $\mathbf{S}$ to $\mathbf{A}$. 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 process be denoted by $\{Y_n, n = 0, 1, \ldots, \infty\}$ and $\{A_n, n = 0, 1, \ldots, \infty\}$. Here, for each $n = 0, 1, \ldots$, 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 $(\mathbf{S}, \mathcal{B}(\mathbf{S}))$, by the Ionescu-Tulcea theorem, c.f., [57, 86], one can construct a probability measure

$\mathbf{P}_\nu^\sigma$ on $(\mathbf{H}_\infty, \mathcal{B}(\mathbf{H}_\infty))$ such that

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

As usual, equalities involving conditional expectations and probabilities are understood in the almost sure sense.

Definition B.1. The probability measure $\mathbf{P}_\nu^\sigma$ is called a strategic measure (of the strategy σ) in the DTMDP model $\{\mathbf{S}, \mathbf{A}, p, \mathrm{l}\}$ (with the initial distribution ν).

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

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

We denote the value function of problem (B.1) by $\mathbf{V}^*$. Then a strategy σ^* is called optimal for problem (B.1) if $\mathbf{V}(x, \sigma^*) = \mathbf{V}^*(x)$ for each $x \in \mathbf{S}$. For a constant $\epsilon > 0$, a strategy is called ϵ-optimal for problem (B.1) if $\mathbf{V}(x, \sigma^*) \leq \mathbf{V}^*(x) + \epsilon$ for each $x \in \mathbf{S}$.

Occasionally we will also consider the so called universally measurable strategies, in which case, the stochastic kernels $\sigma_n(da|h_n)$ are universally measurable, i.e., for each measurable subset Γ of $\mathbf{A}$, $\sigma(\Gamma|h_n)$ is universally measurable in $h_n \in \mathbf{H}_n$. The meaning of universally measurable deterministic Markov or deterministic stationary strategy is understood similarly, i.e., when the underlying mappings are universally measurable in their arguments. See Chapter 7.7 of [9] for the definition of universal measurability and other related measurability concepts, such as the definition of a lower semianalytic function.

We collect the relevant statements in Section 3 of [108] in the next proposition.

Condition B.1. (a) The function $l(x, a, y)$ is lower semicontinuous in $(x, a, y) \in \mathbf{S} \times \mathbf{A} \times \mathbf{S}$.

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

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

Definition B.2. The DTMDP model $\{\mathbf{S}, \mathbf{A}, p, \mathrm{l}\}$ is called semicontinuous if it satisfies Condition B.1.

Condition B.2. (a) The function $l(x, a, y)$ is lower semicontinuous in $a \in \mathbf{A}$ for each $x, y \in \mathbf{S}$.

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

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

Proposition B.1. (a) Let $\mathbf{U}$ be a $[1, \infty]$-valued lower semianalytic function on $\mathbf{S}$. If

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

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

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

$$\mathbf{V}^*(x) = \int_{\mathbf{S}} p(dy|x, \varphi(x)) e^{l(x,\varphi(x),y)} \mathbf{V}^*(y), \ \forall \ x \in \mathbf{S}, \tag{B.2}$$

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

(c) $\mathbf{V}^*(x) = \inf_{\sigma\in\Sigma^U} \mathbf{V}(x,\sigma)$, where Σ^U is the set of universally measurable strategies. Moreover, for each $\epsilon > 0$, there is some universally measurable deterministic stationary ϵ-optimal strategy for problem (B.1).

(d) Suppose Condition B.1 is satisfied. Then the value function $\mathbf{V}^*$ is the minimal $[1,\infty]$-valued lower semicontinuous solution to

$$\mathbf{V}(x) = \inf_{a\in\mathbf{A}} \left\{ \int_{\mathbf{S}} p(dy|x,a) e^{l(x,a,y)} \mathbf{V}(y) \right\}, \ x \in \mathbf{S}. \tag{B.3}$$

(e) Suppose Condition B.2 is satisfied, the value function $\mathbf{V}^*$ is the minimal $[1,\infty]$-valued measurable solution to (B.3).

(f) Suppose Condition B.1 or Condition B.2 is saisfied, let $\mathbf{V}^{(0)}(x) := 1$ for each $x \in \mathbf{S}$, and for each $n = 1, 2, \ldots,$

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

Then $(\mathbf{V}^{(n)}(x))$ increases to $\mathbf{V}^*(x)$ for each $x \in \mathbf{S}$, where $\mathbf{V}^*$ is the value function for problem (B.1). Furthermore, there exists a deterministic stationary strategy φ satisfying (B.2), and so in particular, there exists a deterministic stationary optimal strategy for the risk-sensitive DTMDP problem (B.1).

Part (c) of the above statement follows from the proof of Proposition 3.2 of [108], whereas all the other parts are according to Propositions 3.1, 3.4 and 3.7 therein.