A useful technique for piecewise deterministic Markov decision processes

Guo, Xin; Zhang, Yi

DOI:
10.1016/j.orl.2020.11.002

Document Version
Peer reviewed version

Citation for published version (Harvard):
Guo, X & Zhang, Y 2021, 'A useful technique for piecewise deterministic Markov decision processes', *Operations Research Letters*, vol. 49, no. 1, pp. 55-61. https://doi.org/10.1016/j.orl.2020.11.002

A useful technique for piecewise deterministic Markov decision processes*

Xin Guo † and Yi Zhang ‡

Abstract: This note presents a technique that is useful for the study of piecewise deterministic Markov decision processes (PDMDPs) with general policies and unbounded transition intensities. This technique produces an auxiliary PDMDP from the original one. The auxiliary PDMDP possesses certain desired properties, which may not be possessed by the original PDMDP. We apply this technique to risk-sensitive PDMDPs with total cost criteria, and comment on its connection with the uniformization technique.

Keywords: Continuous-time Markov decision processes. Piecewise deterministic Markov decision processes. Unbounded transition intensities.

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

1 Introduction

This note concerns the optimal control of piecewise deterministic Markov processes, where the state evolves according to a deterministic and uncontrolled flow between two consecutive jumps, and the transition intensities and post-jump distributions are controlled. Below it will be termed as a piecewise deterministic Markov decision process or simply a PDMDP.

A powerful method of studying PDMDPs is to reduce it to an equivalent discrete-time Markov decision process (DTMDP) by inspecting the PDMDP at each of its jump moments and regarding the (possibly relaxed) control function used during a sojourn time as an action in the DTMDP. This method comes back to [20], where it was applied to time non-homogeneous continuous-time Markov decision processes (CTMDPs). (A (homogeneous) CTMDP is a PDMDP, where the state does not change between two consecutive jumps, whereas a time non-homogeneous CTMDP can be viewed as a PDMDP with a specific flow.) Some subsequent applications of this method can be found in e.g., [1, 3, 4, 12, 21]. The action space in the induced DTMDP, as a set of measurable mappings, is in general a more complicated object than the action space in the original PDMDP. The reason for applying this reduction is to gain access to the rich toolbox of known results on DTMDPs that have been studied since 1950s.

It is appreciated that the theory of DTMDPs is better established when the underlying DTMDP model satisfies some compactness-continuity conditions, see [14, 15, 16]. One example of such compactness-continuity conditions is that the action space is a compact Borel space, the loss function

*Declarations of interest: none. External funding body: none.

†School of Economics and Management, Tsinghua University, Beijing, 100084, China. E-mail: guoxin5@sem.tsinghua.edu.cn and x.guooo@hotmail.com. At the time of submission of this paper, Xin Guo was with Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K and the E-mail: X.Guo21@liv.ac.uk.

‡Corresponding author. Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. E-mail: yi.zhang@liv.ac.uk.

is lower semicontinuous in the action, and the transition kernel possesses a strong Feller property with respect to the action, i.e., it maps each bounded measurable function on the state space to a function, which is jointly measurable in the state and action, and also continuous in the action.

However, even if the original PDMDP satisfies a natural set of compactness-continuity conditions, see Condition 3.1 below, it can happen that the transition kernel in the induced DTMDP fails to satisfy the desired continuity condition. We demonstrate this in Example 3.1 below. On the other hand, it turns out that this inconvenience does not appear if the transition intensities of the PDMDP are strongly positive, i.e., bounded away from zero by a constant, see Proposition 3.2.

For CTMDPs with general policies and unbounded transition rates, a different way was proposed in [7] and extended in [9] to reduce the continuous-time problems to DTMDPs, whose action space is the same as the one in the original CTMDP. This method is based on a formula observed in [5], which connects the expected sojourn time with the probability distribution of actions made at the jump epochs. The application of this formula for that purpose is valid if the transition intensities are strongly positive, but could be invalid otherwise, see [7, 13, 17].

The contribution of this paper is that we present a technique, which produces an auxiliary PDMDP model satisfying the following: a) the performance in the original PDMDP can be replicated by a corresponding policy in the auxiliary PDMDP; b) the transition intensities in the auxiliary PDMDP are strongly positive, and its induced DTMDP satisfies the desired compactness-continuity conditions if so does the original PDMDP. Then as an application, we extend some optimality results for risk-sensitive PDMDPs with total cost criteria, which were previously obtained in [12] under the extra requirement on the transition intensities being strongly positive. This requirement is omitted here with the help of the proposed technique. The technique in this note is similar to the technique in [17]. In greater detail, the technique in [17] is based on introducing an additional Poisson process after each jump, whereas here a (controlled) non-homogeneous Poisson process is introduced after each jump. Besides, the technique in [17] produces an auxiliary CTMDP model, which was shown to replicate the (total) occupation measures in the original model. For risk-sensitive problems, the performance measures cannot be readily written as integrals of the cost rate with respect to the occupation measures.

For CTMDPs with uniformly bounded transition intensities, the uniformization technique also produces new CTMDP models, in which the performance measure in the original model can be replicated. It is usually justified under stationary policies. Based on the uniformization technique, one may reduce CTMDPs to equivalent DTMDPs, see e.g [2, 18]. The technique in this paper can be used to justify the reduction method based on uniformitarian for PDMDPs with general policies.

The rest of this paper is organized as follows. In Section 2 we describe the PDMDP model. In Section 3 we present and prove the main statements.

2 Description of PDMDP model

Let $(S, \mathcal{B}(S))$ be a nonempty standard Borel state space, $(A, \mathcal{B}(A))$ be a nonempty standard Borel action space, and q stand for a signed kernel $q(dy|x,a)$ on $\mathcal{B}(S)$ given $(x,a) \in S \times A$ such that $\tilde{q}(\Gamma_S|x,a) := q(\Gamma_S \setminus \{x\}|x,a) \geq 0$ for all $\Gamma_S \in \mathcal{B}(S)$, $q(S|x,a) = 0$, and $\bar{q}_x := \sup_{a\in A} q_x(a) < \infty$, where $q_x(a) := -q(\{x\}|x,a)$ is the transition intensity. The signed kernel q is also called the transition rate. Between two consecutive jumps, the state of the process evolves according to a measurable mapping ϕ from $S \times [0,\infty)$ to S, see (2) below. It is assumed that for each $x \in S$

$$\phi(x, t+s) = \phi(\phi(x,t), s), \ \forall\ s,t \geq 0; \ \phi(x,0) = x, \tag{1}$$

and $t \to \phi(x,t)$ is continuous. Unless stated otherwise, we consider Borel σ-algebras on metric spaces, and the term of measurability is always understood in the Borel sense. Finally let the cost rate c be a $[0,\infty)$-valued measurable function on $S \times A$.

For the rest of this paper, it is convenient to introduce the following notations. Let $\mathbb{P}(A)$ be the space of probability measures on $\mathcal{B}(A)$. For each $\mu \in \mathbb{P}(A)$, we put

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

Condition 2.1 *For each $x \in S$, $\int_0^t \overline{q}_{\phi(x,s)}ds < \infty$, and $\int_0^t \sup_{a\in A} c(\phi(x,s),a)ds < \infty$, for each $t \in [0,\infty)$.*

Condition 2.1 is assumed to hold throughout this paper. The integrals in Condition 2.1 are well defined because the integrands are nonnegative and universally measurable.

Now we briefly describe the PDMDP with the system primitives $\{S, A, q, \phi, c\}$. Let us take the sample space Ω by adjoining to the countable product space $S \times ((0,\infty) \times S)^\infty$ the sequences of the form

$$(x_0, \theta_1, \dots, \theta_n, x_n, \infty, x_\infty, \infty, x_\infty, \dots),$$

where $x_0, x_1, \dots, x_n$ belong to S, $\theta_1, \dots, \theta_n$ belong to $(0,\infty)$, and $x_\infty \notin S$ is the isolated point. We equip Ω with its Borel σ-algebra $\mathcal{F}$.

Let $t_0 := 0 =: \theta_0$, and for each $n \geq 0$, and each element $\omega := (x_0, \theta_1, x_1, \theta_2, \dots) \in \Omega$, let $h_n := (x_0, \theta_1, \dots, \theta_n, x_n)$, $t_n := t_{n-1} + \theta_n$, and $t_\infty(\omega) := \lim_{n\to\infty} t_n$. Then, $(\Omega, \mathcal{F})$ is the canonical sample space of the marked point process (t_n, x_n) with the mark space S, and $\theta_n = t_n - t_{n-1}$ is the sojourn time, where the convention of $\infty - \infty := \infty$ is in use. Define the process, which evolves according to the flow ϕ during a sojourn time:

$$\xi_t = \begin{cases} \phi(x_n, t - t_n), & \text{if } t_n \leq t < t_{n+1}; \\ x_\infty, & \text{if } t_\infty \leq t, \end{cases} \tag{2}$$

where $x_\infty \notin S$ is an isolated cemetery point. The process is controlled through its local characteristics as follows.

A policy π is given by a sequence (π_n) such that, for each $n = 0, 1, 2, \dots$, $\pi_n(da|h_n, s)$ is a stochastic kernel on A given h_n, s with $s > 0$, and for each $\omega = (x_0, \theta_1, x_1, \theta_2, \dots) \in \Omega$, $t > 0$,

$$\begin{aligned} \pi(da|\omega, t) \;:=\; & I\{t \geq t_\infty\}\delta_{a_\infty}(da) \\ & + \sum_{n=0}^{\infty} I\{t_n < t \leq t_{n+1}\}\pi_n(da|h_n, t - t_n), \end{aligned} \tag{3}$$

defines a $\mathbb{P}(A \cup \{a_\infty\})$-valued (relaxed) control process, where $a_\infty \notin A$ is some isolated point. If for some measurable mapping φ from S to A such that $\pi_n(da|x_0, \theta_1, \dots, \theta_n, x, t) \equiv \delta_{\varphi(x)}(da)$, then the policy $\pi = (\pi_n)$ is called deterministic stationary and is identified with the mapping φ.

A policy π and an initial state x define a probability measure P_x^γ on the canonical sample space, under which $P_x^\pi(x_0 = x) = 1$, and the conditional distribution of (θ_{n+1}, x_{n+1}) given h_n satisfies for all

$\Gamma_1 \in \mathcal{B}((0,\infty)),\ \Gamma_2 \in \mathcal{B}(S)$,

$$\begin{aligned} & P^\pi_\gamma(\theta_{n+1} \in \Gamma_1,\ x_{n+1} \in \Gamma_2 | h_n) \\ = & \int_{\Gamma_1} e^{-\int_0^t \int_A q_{\phi(x_n,s)}(a)\pi_n(da|h_n,s)ds} \\ & \times \int_A \tilde{q}(\Gamma_2|\phi(x_n,t),a)\pi_n(da|h_n,t)dt, \\ & P^\pi_\gamma(\theta_{n+1} = \infty,\ x_{n+1} = x_\infty | h_n) \\ = & e^{-\int_0^\infty \int_A q_{\phi(x_n,s)}(a)\pi_n(da|h_n,s)ds} \end{aligned} \tag{4}$$

on $\{x_n \in S\}$.

The proposed technique in this paper will be applied to the risk-sensitive optimal control problem for the PDMDP with a total cost criterion, which is to minimize over all policies π:

$$\begin{aligned} V(x,\pi) &:= E^\pi_x \left[e^{\int_0^\infty \int_A c(\xi_t,a)\pi(da|\omega,t)dt} \right] \\ &= E^\pi_x \left[e^{\sum_{n=0}^\infty \int_{t_n}^{t_{n+1}} \int_A c(\phi(x_n,s-t_n),a)\pi_n(da|h_n,s-t_n)ds} \right]. \end{aligned}$$

Here $\int_{t_n}^{t_{n+1}}$ is understood as $\int_{(t_n,t_n]\bigcap\mathbb{R}}$, and we put $c(x_\infty,a) \equiv 0$. The value function is defined by $V^*(x) = \inf_\pi V(x,\pi)$ for all $x \in S$. We shall call the above system primitives $\{S, A, q, \phi, c\}$ and the corresponding optimal control problem the "original model", to distinguish it from the auxiliary model that will appear later.

We also apply the technique to justifying the uniformization technique for PDMDPs with the expected long run average cost defined by

$$\overline{V}(x,\pi) := \varlimsup_{T\to\infty} E^\pi_x \left[\frac{\int_0^T \int_A c(\xi_t,a)\pi(da|\omega,t)dt}{T} \right].$$

3 Main result

3.1 Observation and auxiliary model

In what follows, let us fix $\lambda_x(a)$ as some measurable function in $(x,a) \in S \times A$ satisfying $\lambda_x(a) \geq \delta > 0$ for all $(x,a) \in S \times A$, $\overline{\lambda}_x := \sup_{a\in A} \lambda_x(a) < \infty$ for all $x \in S$, and $\int_0^t \overline{\lambda}_{\phi(x,s)} ds < \infty$ for each $t \in [0,\infty)$ and $x \in S$.

We introduce an auxiliary model $\{\breve{S}, A, \breve{q}, \breve{\phi}, \breve{c}\}$ defined in terms of the system primitives of the original model as well as the strongly positive function λ. When there is a danger of confusion, we shall primarily use breves to signify the auxiliary model. Without special explanations, all the objects signified with breves are understood similarly to their counterparts without breves.

If $\lambda_x(a) \equiv \lambda > 0$ is a constant, then roughly speaking, the auxiliary model arises from inserting additional inspections of the state process during each sojourn time in the original model (up to the moment of explosion) taking place in an independent Poisson process with rate λ. The changes in the second coordinate of the state in the auxiliary model take place at and only at each of such inspection epochs, which will be recorded as "fictitious" jumps and generate strongly positive transition intensities.

The state space is $\breve{S} = S \times \{-1,1\}$, endowed with the product topology, where $\{-1,1\}$ is with the discrete topology. The action space is A. The transition rate $\breve{q}$ on $\mathcal{B}(\breve{S})$ given $\breve{S} \times A$ is defined as follows: $\breve{q}(dy \times \{-i\}|(x,i),a) = \lambda_x(a)\delta_x(dy)$, $\breve{q}(dy \times \{i\}|(x,i),a) = q(dy|x,a) - \lambda_x(a)\delta_x(dy)$ with $\delta_x(dy)$ being the Dirac measure concentrated on the singleton $\{x\}$, so that $\breve{q}_{(x,i)}(a) = \breve{q}(\breve{S} \setminus \{(x,i)\}|(x,i),a) =$

$\breve{q}(S \setminus \{x\} \times \{-1,1\}|(x,i),a) + \breve{q}(S \times \{-i\}|(x,i),a) = q_x(a) + \lambda_x(a)$ for all $(x,i) \in \breve{S},\ a \in A$. In other words, the auxiliary model has strongly positive transition intensities. The flow is defined by $\breve{\phi}((x,i),t) = (\phi(x,t),i)$. The cost rate is $\breve{c}((x,i),a) = c(x,a)\ \forall\ (x,i) \in \breve{S},\ a \in A$. Let $\breve{V}((x,i),\breve{\pi}) = \breve{E}^{\breve{\pi}}_{(x,i)}[e^{\int_0^\infty \breve{c}(\breve{\xi}_t,a)\breve{\pi}(da|\breve{\omega},t)dt}]$.

Definition 3.1 *Consider the canonical sample space of the marked point process $(\breve{t}_n, x_n, i_n)$, and a sample path $\breve{\omega} = ((x_0,i_0),\breve{\theta}_1,(x_1,i_1),\dots,\breve{\theta}_n,(x_n,i_n),\dots)$. We say a mark (x_l,i_l) $(l \geq 1)$ is immediately after a fictitious jump if $i_l = -i_{l-1}$, or equivalently, $x_l = \phi(x_{l-1},\breve{\theta}_l)$, where $\breve{\theta}_l$ is the sojourn time before the mark (x_l,i_l). A mark that is not immediately after a fictitious jump is called immediately after an honest jump. We regard (x_0,i_0) as a mark immediately after an honest jump.*

Using the notation in the above definition, we may consider out of $(\breve{t}_n, x_n, i_n)$ another process $(\tau_{(m)}, x_{(m)}, i_{(m)})$ with $\tau_{(0)} := 0$ by counting only the points with marks immediately after honest jumps. If $\tau_{(m)} = \infty$ for some m, then we put $\tau_{(m+1)} = \infty$ and $x_{(m+1)} = (x_\infty, i_{(0)})$. Since (x_0,i_0) is regarded as a mark immediately after an honest jump, $x_0 = x_{(0)}$ and $i_0 = i_{(0)}$. Since $i_{(m)} = i_{(0)}$ for all $m \geq 0$ almost surely in $(\tau_{(m)}, x_{(m)}, i_{(m)})$, with $i_{(0)}$ being fixed we may simply consider the marked point process $(\tau_{(m)}, x_{(m)})$ instead of $(\tau_{(m)}, x_{(m)}, i_{(m)})$.

Part (a) of the next statement is a generalization of Theorem 2.1 of [19].

Theorem 3.1 *Suppose Condition 2.1 is satisfied. For each policy $\pi = (\pi_n)$ in the original PDMDP model, there is a policy $\breve{\pi} = (\breve{\pi}_n)$ in the auxiliary PDMDP model such that for all $x \in S$ and $i \in \{-1,1\}$:*

(a) The distribution of the marked point process $(\tau_{(m)}, x_{(m)})$ under $\breve{P}^{\breve{\pi}}_{(x,i)}$ coincides with the distribution of the process (t_m, x_m) under P^π_x.

(b) $V(x,\pi) = \breve{V}((x,i),\tilde{\pi})$ and $\overline{V}(x,\pi) = \breve{\overline{V}}((x,i),\breve{\pi})$.

Proof. We will make use of the notation in Definition 3.1 freely.

(a) Let a policy $\pi = (\pi_n)$ for the original model be fixed. Consider the corresponding policy $\breve{\pi} = (\breve{\pi}_n)$ in the auxiliary model defined as follows. For the n-history

$$\breve{h}_n = ((x_0,i_0),\breve{\theta}_1,(x_1,i_1),\dots,(x_{n-1},i_{n-1}),\breve{\theta}_n,(x_n,i_n))$$

in the auxiliary model, let $m = m(\breve{h}_n)$ be the number of honest jumps over $(0,\breve{t}_n]$ within $\breve{h}_n$, so that if we count the initial mark (i_0,x_0) as immediately after an honest jump, then there are $m+1$ marks immediately after honest jumps within $\breve{h}_n$. Then we define

$$\begin{aligned}\breve{\pi}_n(da|\breve{h}_n,t) &= \pi_m(da|x_0,\tau_{(1)},x_{(1)},\tau_{(2)}-\tau_{(1)},\dots,\\ &\tau_{(m)}-\tau_{(m-1)},x_{(m)},t+\breve{t}_n-\tau_{(m)})\ \forall\ t>0. \end{aligned} \tag{5}$$

Consequently, for each $n,m \geq 0$ and for each $t \in (0,\infty)$ satisfying $t \in (\breve{t}_n,\breve{t}_{n+1}] \subseteq (\tau_{(m)},\tau_{(m+1)}]$, we have

$$\begin{aligned}\breve{\pi}(da|\breve{\omega},t) &= \breve{\pi}_n(da|\breve{h}_n,t-\breve{t}_n) = \pi_m(da|x_0,\tau_{(1)},x_{(1)},\\ &\tau_{(2)}-\tau_{(1)},\dots,\tau_{(m)}-\tau_{(m-1)},x_{(m)},t-\tau_{(m)}),\end{aligned} \tag{6}$$

where the first equality is by (3) applied to $\breve{\pi}$.

For brevity, below we put

$$\begin{aligned}\tilde{q}(dy|\phi(x_{(m)},t),\pi_m) &:= \int_A \tilde{q}(dy|\phi(x_{(m)},t),a)\\ &\pi_m(da|x_{(0)},\tau_{(1)},x_{(1)},\tau_{(2)}-\tau_{(1)},\dots,x_{(m)},t),\\ q(dy|\phi(x_{(m)},t),\pi_m) &:= \int_A q(dy|\phi(x_{(m)},t),a)\\ &\pi_m(da|x_{(0)},\tau_{(1)},x_{(1)},\tau_{(2)}-\tau_{(1)},\dots,x_{(m)},t),\end{aligned}$$

and $q_{\phi(x_{(m)},t)}(\pi_m) := \tilde{q}(S|\phi(x_{(m)},t),\pi_m)$. We similarly understand the notation $\lambda_{\phi(x_{(m)},t)}(\pi_m)$, $(\lambda+q)_{\phi(x_{(m)},t)}(\pi_m)$ and $c(\phi(x_{(m)},t),\pi_m)$.

Now let us show that the distribution of the marked point process (τ_m, x_m) under $\breve{P}^{\breve{\pi}}_{(x,i_0)}$ coincides with the distribution of the marked point process in the original model under P^{π}_{x}. To this end, in view of (4), $x_{(0)} = x_0$ and $\tau_{(0)} = 0$, it is sufficient to show that

$$\begin{aligned}
&\breve{P}^{\breve{\pi}}_{(x,i)}(x_{(m+1)} \in \Gamma,\ \tau_{(m+1)} - \tau_{(m)} \in [0,T] | x_{(0)}, \tau_{(1)}, \\
&x_{(1)}, \ldots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}) \\
= &\int_0^T \tilde{q}(\Gamma|\phi(x_{(m)},t),\pi_m) e^{-\int_0^t q_{\phi(x_{(m)},s)}(\pi_m) ds} dt
\end{aligned} \tag{7}$$

on $\{\tau_{(m)} < \infty\}$ for each $T > 0$, $\Gamma \in \mathcal{B}(S)$ and $m \geq 0$. Equality (7) would be justified once we show that

$$\begin{aligned}
&\breve{P}^{\breve{\pi}}_{(x,i)}(x_{(m+1)} \in \Gamma,\ \tau_{(m+1)} - \tau_{(m)} \in [0,T], \\
&\text{ exactly } n \text{ ficticious jumps over } [\tau_{(m)}, \tau_{(m+1)}] |\ x_{(0)}, \\
&\tau_{(1)}, x_{(1)}, \ldots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}) \\
= &\int_0^T \frac{(\int_0^{v_{n+1}} \lambda_{\phi(x_{(m)},s)}(\pi_m) ds)^n}{n!} \tilde{q}(\Gamma|\phi(x_{(m)}, v_{n+1}), \pi_m) \\
&\times e^{-\int_0^{v_{n+1}} (\lambda+q)_{\phi(x_m,s)}(\pi_m) ds} dv_{n+1}.
\end{aligned}$$

Indeed, the expression on the left-hand side of the previous equality can be written as

$$\begin{aligned}
&\int_0^T \int_0^{T-r_1} \cdots \int_0^{T-\sum_{i=1}^n r_i} \tilde{q}(\Gamma|\phi(x_{(m)}, \sum_{i=1}^n r_i + t), \pi_m) \\
&\times \left(\prod_{j=1}^n \lambda_{\phi(x_{(m)}, \sum_{i=1}^j r_i)}(\pi_m) \right) \\
&\times e^{-\int_0^{\sum_{i=1}^n r_i + t} (\lambda+q)_{\phi(x_{(m)},s)}(\pi_m) ds} dt dr_{n-1} \ldots dr_1
\end{aligned}$$

With the change of variables: $r_1 \to v_1$, $r_1 + r_2 \to v_2, \ldots, \sum_{i=1}^n r_i \to v_n, \sum_{i=1}^n r_i + t \to v_{n+1}$, the previous integral can be written as

$$\begin{aligned}
&\int_0^T \int_{v_1}^T \cdots \int_{v_n}^T \tilde{q}(\Gamma|\phi(x_{(m)}, v_{n+1}), \pi_m) \\
&\times \left(\prod_{j=1}^n \lambda_{\phi(x_{(m)}, v_j)}(\pi_m) \right) \\
&\times e^{-\int_0^{v_{n+1}} (\lambda+q)_{\phi(x_{(m)},s)}(\pi_m) ds} dv_{n+1} \ldots dv_2 dv_1,
\end{aligned}$$

which, by the Fubini theorem, coincides with

$$\begin{aligned}
& \int_0^T \int_0^{v_{n+1}} \int_{v_1}^{v_{n+1}} \cdots \int_{v_{n-1}}^{v_{n+1}} \prod_{j=1}^{n} \lambda_{\phi(x_{(m)},v_j)}(\pi_m) \\
& dv_n dv_{n-1} \ldots dv_2 dv_1 \ \tilde{q}(\Gamma|\phi(x_{(m)}, v_{n+1}), \pi_m) \\
& \times e^{-\int_0^{v_{n+1}} (\lambda+q)_{\phi(x_{(m)},s)}(\pi_m) ds} dv_{n+1} \\
= & \int_0^T \int_{\{0\le v_1\le v_2\le\cdots\le v_n\le v_{n+1}\}} \prod_{j=1}^{n} \lambda_{\phi(x_{(m)},v_j)}(\pi_m) \\
& dv_1 dv_2 \ldots dv_n \ \tilde{q}(\Gamma|\phi(x_{(m)}, v_{n+1}), \pi_m) \\
& \times e^{-\int_0^{v_{n+1}} (\lambda+q)_{\phi(x_{(m)},s)}(\pi_m)) ds} dv_{n+1} \\
= & \int_0^T \frac{(\int_0^{v_{n+1}} \lambda_{\phi(x_{(m)},s)}(\pi_m) ds)^n}{n!} \tilde{q}(\Gamma|\phi(x_{(m)}, v_{n+1}), \pi_m) \\
& \times e^{-\int_0^{v_{n+1}} (\lambda+q)_{\phi(x_m,s)}(\pi_m) ds} dv_{n+1},
\end{aligned}$$

as required, where for the last equality, one may either recognize it as a known fact, or more directly, recall that if $Z = (X_{(1)}, \ldots, X_{(n)})$ is the order statistic of i.i.d. $[0,\infty)$-valued continuous random variables $X_1, \ldots, X_n$ with the common marginal p.d.f. f, then $n! \prod_{i=1}^n f(t_i)$ for $t_1 \le t_2 \le \cdots \le t_n$ defines the joint density of Z, so that

$$\begin{aligned}
& \int_{\{0\le t_1\le t_2\le\cdots\le t_n\le t\}} n! \prod_{i=1}^{n} f(t_i) dt_1 \ldots dt_n \\
= & \ P(X_{(n)} \le t) = P(X_1 \le t)^n = \left(\int_0^t f(s) ds \right)^n .
\end{aligned}$$

Part (a) is thus proved.

(b) It follows that

$$\begin{aligned}
& \breve{V}((x,i_0),\breve{\pi}) = \breve{E}^{\breve{\pi}}_{(x,i_0)}[e^{\int_0^\infty \breve{c}(\breve{\xi}_t,a)\breve{\pi}(da|\breve{\omega},t)dt}] \\
= & \ \breve{E}^{\breve{\pi}}_{(x,i_0)} \left[e^{\sum_{m=0}^\infty \int_{\tau_{(m)}}^{\tau_{(m+1)}} c(\phi(x_{(m)}, s-\tau_{(m)}),\pi_m),\pi_m) ds} \right] \\
= & \ V(x,\pi),
\end{aligned}$$

where the second equality holds by the assumed local integrability of $\overline{\lambda}_{\phi(x,s)}$ in $s > 0$, (6) and the definition of $\breve{c}$, and the last equality follows from part (a). The last assertion follows similarly from Theorem 3.1(a) and

$$\begin{aligned}
\overline{V}(x,\pi) = & \varlimsup_{T\to\infty} \frac{1}{T} E^\pi_x \left[\sum_{n=0}^{\infty} \int_{t_n\wedge T}^{t_{n+1}\wedge T} \right. \\
& \left. \times \int_A c(\phi(x_n, s-t_n), a) \pi_n(da|h_n, s-t_n) ds \right],
\end{aligned}$$

where $t_n \wedge T := \min\{t_n, T\}$. □

Remark 3.1 *(a) By Theorem 3.1(b), $\breve{V}^*(x) \le V^*(x)$ for each $x \in S$.*
(b) By inspecting the proof of Theorem 3.1 (see especially (5) and (6) therein), one can tell that for a deterministic stationary policy in the auxiliary model, which depends on $(x,i) \in S \times \{-1,1\}$ only

through $x \in S$, and is identified by a measurable mapping φ from S to A, $\breve{V}((x,i),\varphi) = V(x,\varphi)$ for all $x \in S$ and $i \in \{-1,1\}$. Therefore, if such a deterministic stationary policy φ is optimal in the auxiliary model, then so is it in the original model, and $V^*(x) = \breve{V}^*(x) = \breve{V}((x,i),\varphi) = V(x,\varphi)$ for each $x \in S$.

3.2 Application to risk-sensitive PDMDPs

In this subsection, S and A are topological Borel spaces. A topological Borel space is a topological space that is homeomorphic to a Borel subset (endowed with the relative topology) of a Polish space. Let $\mathcal{B}(S)$ and $\mathcal{B}(A)$ be the Borel σ-algebras on S and A. We endow $\mathbb{P}(A)$ with the standard weak topology. The results in the previous subsection are all applicable.

Let us introduce a natural set of compactness-continuity conditions on the original PDMDP.

Condition 3.1 *(a) For each bounded measurable function f on S and each $x \in S$, $\int_S f(y)\tilde{q}(dy|x,a)$ is continuous in $a \in A$.*
(b) For each $x \in S$, the (nonnegative) function $c(x,a)$ is lower semicontinuous in $a \in A$.
(c) The action space A is a compact metric space.

The usefulness of the auxiliary PDMDP also partially lies in the next observation.

Lemma 3.1 *Let $\lambda_x(a) \equiv \lambda > 0$ be a constant. If the original PDMDP model satisfies Conditions 2.1 and 3.1, then the auxiliary model satisfies the corresponding versions of Conditions 2.1 and 3.1, too.*

Proof. We only verify the version of Condition 3.1(a). For any bounded measurable function f on $\breve{S}$, it holds that

$$\begin{aligned}
& \int_{\breve{S}} f(y,j)\tilde{\breve{q}}(d(y,j)|(x,i),a) \\
= & \int_{\breve{S}} f(y,j)\breve{q}(d(y,j)|(x,i),a) + f(x,i)\breve{q}_{(x,i)}(a) \\
= & \int_S f(y,i)(q(dy|x,a) - \lambda\delta_x(dy)) + \int_S f(y,-i)\lambda\delta_x(dy) \\
& + f(x,i)(\lambda + q_x(a)) \\
= & \int_S f(y,i)\tilde{q}(dy|x,a) - q_x(a)f(x,i) - \lambda f(x,i) \\
& + \lambda f(x,-i) + f(x,i)(\lambda + q_x(a)) \\
= & \int_S f(y,i)\tilde{q}(dy|x,a) + \lambda f(x,-i),
\end{aligned}$$

which is clearly continuous in $a \in A$ when the original model satisfies Condition 3.1. □

The following statement was obtained in Theorem 3.1 and Remark 3.1 of [12].

Proposition 3.1 *Suppose Conditions 2.1 and 3.1 are satisfied. In addition, $\inf_{(x,a)\in S\times A} q_x(a) > 0$. (See the discussions below Proposition 3.2 regarding this additional assumption.) Then the following assertions hold.*

(a) The value function V^* is the minimal $[1,\infty]$-valued measurable solution to the following opti-

mality equation:

$$
\begin{aligned}
& -(V(\phi(x,t)) - V(x)) \\
= \; & \int_0^t \inf_{a\in A} \Big\{ \int_S V(y)\tilde{q}(dy|\phi(x,\tau),a) \\
& -(q_{\phi(x,\tau)}(a) - c(\phi(x,\tau),a))V(\phi(x,\tau)) \Big\}\, d\tau \\
& \forall\ t \in [0,\infty), x \in S^*; \\
& V(x) < \infty\ \forall\ x \in S^*;\ V(x) = \infty\ \forall\ x \notin S^*
\end{aligned}
\tag{8}
$$

with $S^* := \{x \in S :\ V^*(x) < \infty\}$. *In particular,* $V^*(\phi(x,t))$ *is absolutely continuous in* t *for each* $x \in S^*$.

(b) Any measurable mapping φ *from* S *to* A *such that*

$$
\begin{aligned}
& \inf_{a\in A} \Big\{ \int_S V^*(y)\tilde{q}(dy|x,a) \\
& -(q_x(a) - c(x,a))V^*(x) \Big\} \\
= \; & \int_S V^*(y)\tilde{q}(dy|x,\varphi(x)) \\
& -(q_x(\varphi(x)) - c(x,\varphi(x)))V^*(x),\ \forall\ x \in S^*.
\end{aligned}
$$

defines a deterministic stationary optimal policy in the original model. Such measurable selectors φ *exist.*

The proof of Proposition 3.1 in [12] is based on the study of a DTMDP model induced by the PDMDP model by inspecting the PDMDP at each of its jump moments and regarding the relaxed control functions used during a sojourn time as the actions in the DTMDP. The first coordinate in the state space of the induced DTMDP records the most recent sojourn time, and the second coordinate records the state in the PDMDP immediately after the corresponding jump. More precisely, the DTMDP induced by the PDMDP $\{S, A, q, \phi, c\}$ is specified by the following system primitives:

- The state space is $\mathbf{X} := ((0,\infty) \times S) \cup \{(\infty, x_\infty)\}$. Whenever the topology is concerned, (∞, x_∞) is regarded as an isolated point in $\mathbf{X}$.

- The action space is $\mathbf{A} := \mathcal{R}$, where $\mathcal{R}$ is the space of $\mathbb{P}(A)$-valued measurable mappings $\rho = (\rho_t(da))$ on $(0,\infty)$. Two elements in $\mathcal{R}$ that coincide almost everywhere are not distinguished. We endow $\mathcal{R}$ with the Young topology, which is the weakest topology with respect to which the function $\int_0^\infty \int_A f(t,a)\rho_t(da)dt$ is continuous in $\rho \in \mathcal{R}$ for each strongly integrable Carathéodory function f on $(0,\infty) \times A$. Here a real-valued measurable function f on $(0,\infty) \times A$ is called a strongly integrable Carathéodory function if for each fixed $t \in (0,\infty)$, $f(t,a)$ is continuous in $a \in A$, and for each fixed $a \in A$, $\sup_{a\in A} |f(t,a)|$ is integrable in t, i.e., $\int_0^\infty \sup_{a\in A} |f(t,a)|dt < \infty$. According to Section 43 of [4], see Proposition 43.3 therein, if A is a compact metric space, then so is $\mathcal{R}$ (with a compatible metric).

- The transition kernel p on $\mathcal{B}(\mathbf{X})$ from $\mathbf{X} \times \mathbf{A}$ is given for each $\rho = (\rho_t(da))_{t>0} \in \mathbf{A}$ by

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

$p(\{(\infty, x_\infty)\}|(\theta, x), \rho) := e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds}$ for all $x \in S$, $\theta \in (0, \infty)$, $p(\{(\infty, x_\infty)\}|(\infty, x_\infty), \rho) := 1$. (Recall that the notation $q(dy|x, \rho_t) = \int_A q(dy|x, a)\rho_t(da)$ is in use.)

- The cost function l is a $[0, \infty]$-valued measurable function on $\mathbf{X}\times\mathbf{A}\times\mathbf{X}$ given for all $((\theta, x), \rho, (\tau, y)) \in \mathbf{X} \times \mathbf{A} \times \mathbf{X}$ by

$$\begin{aligned} & l((\theta, x), \rho, (\tau, y)) \\ := & \int_0^\infty I\{s < \tau\}c(\phi(x, s), \rho_s)ds. \end{aligned}$$

For the induced DTMDP $\{\mathbf{X}, \mathbf{A}, p, l\}$, following the reasoning in the proof of Lemma 3.2 of [3] and Chapter 4 of [4], one can see the following statement, which is important for the reasoning in [12].

Proposition 3.2 *Under Conditions 2.1 and 3.1, for each $(\theta, x) \in \boldsymbol{X}$ and $(\tau, y) \in \boldsymbol{X}$, $\rho \in \boldsymbol{A} \to l((\theta, x), \rho, (\tau, y))$ is lower semicontinuous, and $\boldsymbol{A}$ is a compact metric space. If in addition, the transition intensities are strongly positive, then for each $(\theta, x) \in \boldsymbol{X}$, $\rho \in \boldsymbol{A} \to \int_{\boldsymbol{X}} f(z)p(dz|(\theta, x), \rho)$ is continuous for each bounded measurable function f on $\boldsymbol{X}$.*

The strong positivity requirement on the transition intensities is important to the correctness of the previous proposition. This requirement was unfortunately missing and overlooked in [12]. Indeed, the proof of Lemma 4.1 of [12] made use of the strong Feller property of the transition kernel p in the induced DTMDP, which could fail to hold without this additional requirement, as demonstrated in Example 3.1. More precisely, if the transition intensities are not strongly positive, then it can happen that $\int_{\mathbf{X}} f(z)p(dz|(\theta, x), \rho)$ is not continuous for some bounded measurable function f on $\mathbf{X}$.

Example 3.1 *Suppose S is any finite set (endowed with discrete topology), and $A = [0, 1]$, which is a compact metric space, $q_x(a) = a$ and $c(x, a) \equiv 0$, and $\phi(x, t) \equiv x$. Evidently, Conditions 2.1 and 3.1 are satisfied by this PDMDP model. Consider $\rho \in \boldsymbol{A}$ and $(\rho^{(n)}) \subseteq \boldsymbol{A}$ defined as follows: 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 Carathéodory 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 $n \to \infty$. (Recall that $\boldsymbol{A}$ is endowed with the Young topology.) Now for $f(t, x) \equiv 0$ on $(0, \infty) \times S$ and $f(\infty, x_\infty) = 1$, $\int_{\boldsymbol{X}} f(z)p(dz|(\theta, x), \rho^{(n)}) = e^{-\int_0^\infty q_x(\rho_s^{(n)})ds} = e^{-\int_0^\infty \frac{1}{n}ds} = 0 < 1 = e^{-\int_0^\infty 0ds} = \int_{\boldsymbol{X}} f(z)p(dz|(\theta, x), \rho)$.*

As an application of Theorem 3.1 (more precisely, Remark 3.1 drawn from it), we may remove the redundant condition on the strong positivity of the transition intensities from Proposition 3.1.

Corollary 3.1 *Under Conditions 2.1 and 3.1, the assertions stated in Proposition 3.1 all hold without the requirement* $\inf_{(x,a)\in S\times A} q_x(a) > 0$.

Proof. The statement follows from Remark 3.1 and applying Proposition 3.1 to the auxiliary model with a constant $\lambda_x(a) \equiv \lambda > 0$, which is legitimate in view of Lemma 3.1 and that the transition intensities in the auxiliary model are strongly positive. The details are as follows.

Step 1. We show that the value function $\breve{V}^*((x, i))$ in the auxiliary PDMDP model depends on (x, i) only through x, and can thus be identified as $\breve{V}^*(x)$.

For this, we will apply the following result from [12]: under the conditions in Proposition 3.1, including that the transition intensities are strongly positive:

- The value function V^* in the original model is the minimal $[1,\infty]$-valued measurable solution to the optimality equation $V = \mathcal{T} \circ V$, where for all $x \in S$

$$\begin{aligned} & \mathcal{T} \circ V(x) \\ := \quad & \inf_{\rho \in \mathcal{R}} \left\{ \int_0^\infty e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s),\rho_s))ds} \right. \\ & \left(\int_S V(y) \tilde{q}(dy|\phi(x,\tau),\rho_\tau) \right) d\tau \\ & \left. + e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds} e^{\int_0^\infty c(\phi(x,s),\rho_s)ds} \right\}. \end{aligned} \tag{9}$$

  Here and below, $e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds} e^{\int_0^\infty c(\phi(x,s),\rho_s)ds} := 0$ whenever $e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s)ds} = 0$.

- The value function V^* can be obtained from the successive approximation: $V^*(x) = \lim_{n\to\infty} \mathcal{T} \circ V_0(x)$ with $V_0(x) \equiv 1$.

According to Lemma 3.1 and that the transition intensities in the auxiliary model are strongly positive (minorized by λ), we may apply the result just quoted above to the auxiliary model and conclude that $\breve{V}^*$ is the minimal $[1,\infty]$-valued measurable function to the following equation

$$\begin{aligned} & \breve{V}((x,i)) = \inf_{\rho \in \mathcal{R}} \left\{ \int_0^\infty e^{\int_0^\theta \breve{c}(\breve{\phi}((x,i),s),\rho_s)ds} \right. \\ & \times e^{-\int_0^\theta \breve{q}_{\breve{\phi}((x,i),s)}(\rho_s)ds} \\ & \left. \times \left(\int_{\breve{S}} \breve{V}((y,j)) \tilde{\breve{q}}(d(y,j)|\breve{\phi}((x,i),\theta),\rho_\theta) \right) d\theta \right\} \\ = \quad & \inf_{\rho \in \mathcal{R}} \left\{ \int_0^\infty e^{\int_0^\theta c(\phi(x,s),\rho_s)ds} e^{-\int_0^\theta (q_{\phi(x,s)}(\rho_s) + \lambda)ds} \right. \\ & \times \left(\int_S \tilde{q}(dy|\phi(x,\theta),\rho_\theta) \breve{V}((y,i)) \right. \\ & \left. \left. + \lambda \breve{V}((\phi(x,\theta),-i)) \right) d\theta \right\}. \end{aligned}$$

Moreover, $\breve{V}^*$ is the pointwise limit of the sequence of functions $\{\breve{V}_n\}_{n=0}^\infty$ with

$$\begin{aligned} & \breve{V}_0((x,i)) :\equiv 1, \\ & \breve{V}_{n+1}((x,i)) := \inf_{\rho \in \mathcal{R}} \left\{ \int_0^\infty e^{\int_0^\theta c(\phi(x,s),\rho_s)ds} \right. \\ & \times e^{-\int_0^\theta (q_{\phi(x,s)(\rho_s)} + \lambda)ds} \left(\int_S \tilde{q}(dy|\phi(x,\theta),\rho_\theta) \breve{V}_n((y,i)) \right. \\ & \left. \left. + \lambda \breve{V}_n((\phi(x,\theta),-i)) \right) d\theta \right\}. \end{aligned}$$

An inductive argument reveals that $\breve{V}_{n+1}((x,i))$ does not depend on i for all $n \geq 0$ and thus $\breve{V}^*((x,i))$ does not depend on i. Below, we write $\breve{V}^*(x)$ for $\breve{V}^*((x,i))$.

Step 2. Again by Lemma 3.1 and that the transition intensities in the auxiliary model are strongly positive, we apply Proposition 3.1(b) to the auxiliary model to obtain a deterministic stationary optimal policy φ. It is possible to take φ, which only depends on $x \in S$ (independent on $i \in \{-1,1\}$)

because for each $(x,i) \in \breve{S}^* := \{x \in S : \breve{V}^*(x) < \infty\} \times \{-1,1\}$,

$$\begin{aligned} & \int_{\breve{S}} \breve{V}^*(y)\tilde{\breve{q}}(d(y,j)|(x,i),a) - (\breve{q}_{(x,i)}(a) \\ & -\breve{c}((x,i),a))\breve{V}^*(x)) \\ = & \int_S \breve{V}^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c(x,a))\breve{V}^*(x)) \end{aligned}$$

does not involve $i \in \{-1,1\}$, where the equality holds by the definition of $\breve{q}$ and $\breve{c}$ and a similar calculation as the one in the proof of Lemma 3.1.

Step 3. Step 2 and Remark 3.1(b) imply $\breve{V}^*(x) = V^*(x)$. Note that the optimality equations for both the original model and the auxiliary model are the same and given by (8), the statement of this corollary follows from applying Proposition 3.1 to the auxiliary model again. □

We end this subsection with the following remarks.

Remark 3.2 *(a) One may modify Condition 3.1 by requiring $\phi(x,t)$ to be continuous in (x,t), and replacing its (a) by (a'): For each bounded continuous function f on S, $(x,a) \in S \times A \to \int_S f(y)\tilde{q}(dy|x,a)$ is continuous; and (b) by (b'): $c(x,a)$ is lower semicontinuous in $(x,a) \in S \times A$. Under the resulting condition, which is often termed as (W), a version of the assertions of Proposition 3.1 and Corollary 3.1 still holds, where V^* is the minimal $[1,\infty]$-valued lower semicontinuous solution to (3.1). A further extension of (W) would be to allow the set of admissible actions depending on the current state, and instead of (b') and the compactness of A, one requires the cost function to be $\mathbb{K}$-inf-compact. The concept of $\mathbb{K}$-inf-compactness was introduced in [10], see also [11]. We do not consider this extension here for simplicity.*
(b) The problem of $V(x,\pi) \to \min_\pi$ is for a risk-averse controller; one may consider the problem

$$W(x,\pi) := E_x^\pi \left[e^{-\int_0^\infty \int_A c(\xi_t,a)\pi(da|\omega,t)dt} \right] \to \max_\pi$$

for a risk-seeking controller. Let its value function be denoted by W^*. Then suitable versions of Proposition 3.1 and Corollary 3.1 hold for this risk-seeking problem. E.g., W^* is a solution to

$$\begin{aligned} & -(V(\phi(x,t)) - V(x)) \\ = & \int_0^t \sup_{a \in A} \Big\{ \int_S V(y)\tilde{q}(dy|\phi(x,\tau),a) \\ & -(q_{\phi(x,\tau)}(a) + c(\phi(x,\tau),a))V(\phi(x,\tau)) \Big\} d\tau \\ & \forall\ t \in [0,\infty), x \in S, \end{aligned}$$

and a deterministic stationary optimal policy exists. Again, a suitable version also holds when Condition 3.1 is replaced by (W), in which case, W^* will be upper semicontinuous.
(c) Consider the DTMDP $\{\breve{\boldsymbol{X}}, \boldsymbol{A}, \breve{p}, \breve{l}\}$, which is the same as the one defined below Proposition 3.1 except that it is now induced by the auxiliary model instead of the original model. The state in $\breve{\boldsymbol{X}}$ is in the form (θ,x,i), but one may get rid of the coordinate i from the state, since it is inessential information, see [8]. The performance of any policy in the original DTMDP can be reproduced by a policy in the DTMDP with i removed from the state. The transition probability in the new DTMDP is given by $e^{-\int_0^t (\lambda+q)_{\phi(x,s)}(\rho_s)ds}(\lambda_{\phi(x,t)}(\rho_t)\delta_{\phi(x,t)}(dy) + \tilde{q}(dy|\phi(x,t),\rho_t))dt$. This DTMDP also satisfies the desired properties similar to those described in Proposition 3.2.

3.3 Connection with uniformization techinque

Uniformization technique is useful in reducing a CTMDP with a uniformly bounded transition rate to an equivalent DTMDP problem. We may use the observation in Subsection 3.1 to justify this

technique for PDMDPs with the average criterion $\overline{V}(x,\pi)$. Assume that $\sup_{x\in S}\overline{q}_x < K$ for a constant $K>0$. For simplicity we assume that the cost rate c is bounded. For the auxiliary model, we put $\lambda_x(a) = K - q_x(a)$ for all $(x,a)\in S$ so that $\breve{q}_x(a)\equiv K$. Theorem 3.1(b) asserts that for any policy π, there is some policy $\breve{\pi}$ in the auxiliary model such that $\overline{V}(x,\pi) = \breve{\overline{V}}((x,i),\breve{\pi})$. The reasoning in the proof of Lemma 2.2 of [6] shows that

$$\begin{aligned}\breve{\overline{V}}(x,i) &= \varlimsup_{N\to\infty}\frac{1}{N}\breve{E}_{(x,i)}\left[\sum_{n=0}^{N-1}K\int_0^\infty e^{-Ks}\right.\\ &\quad\left.\times\int_A c(\phi(x_n,s),a)\breve{\pi}_n(da|\breve{h}_n,s)ds\right]\end{aligned}$$

The above expression is the long run average cost in the DTMDP $\{\breve{\mathbf{X}},\mathbf{A},\breve{p},\breve{\breve{l}}\}$, which is the same as the one introduced below Proposition 3.1 except for $\breve{\breve{l}}((\theta,x,i),\rho,(\tau,y,j)) := K\int_0^\infty e^{-Ks}c(\phi(x,s),\rho_s)ds =: \breve{\breve{l}}(x,\rho)$, and that it is induced by the auxiliary model instead of the original model, hence justifying the different notation. Any policy $\breve{\pi}$ in the auxiliary model is identified with a policy in this DTMDP. Thus, we reduce the PDMDP problem with average criterion to a similar DTMDP problem. Moreover, according to Theorem 2 of [8], we may get rid of the inessential information (θ,i) from the state (θ,x,i) and the original problem is reduced to a similar problem for a simpler DTMDP with state space S, action space $\mathbf{A}$, the transition probability given by

$$\begin{aligned}Q(dy|x,\rho) &= \int_0^\infty e^{-Kt}(K-q_{\phi(x,t)}(\rho_t))\delta_{\phi(x,t)}(dy)\\ &\quad +\tilde{q}(dy|\phi(x,t),\rho_t))dt\end{aligned}$$

and the cost function $\breve{\breve{l}}(x,\rho)$. Any deterministic stationary optimal policy for this DTMDP problem produces an optimal policy in the original PDMDP problem.

Acknowledgements

We thank Dr Yonghui Huang for drawing our attention on the error in [12] mentioned below Proposition 3.2, and the referee and the area editor (Prof. Eugene Feinberg) for insightful remarks.

References

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

[2] Beutler, F. and Ross, K. (1987). Uniformization for semi-Markov decision processes under stationary policies. *J. Appl. Probab.* **24**, 644-656.

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

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

[5] Feinberg, E.A. (1994). A generalization of 'expectation equals reciprocal of intensity' to non-stationary exponential distributions. *J. Appl. Probab.* **31**, 262–267.

[6] Feinberg, E.A. (2002). Optimal control of average reward constrained continuous-time finite Markov decision processes. *Proc. of the 41st IEEE CDC, Las Vegas*, 3805–3810.

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

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

[9] Feinberg, E.A. (2012). Reduction of discounted continuous-time MDPs with unbounded jump and reward rates to discrete-time total-reward MDPs. In *Optimization, Control, and Applications of Stochastic Systems*, Hernandez-Hernandez, D. and Minjarez-Sosa, A. (eds): 77–97, Birkhäuser, Bassel.

[10] Feinberg, E.A., Kasyanov, P. and Zadoianchuk, N. (2012). Average cost Markov decision processes with weakly continuous transition probabilities. *Math. Oper. Res.* **37**, 591–607.

[11] Feinberg, E.A., Kasyanov, P. and Zadoianchuk, N. (2013). Berge's theorem for noncompact image sets. *J. Math. Anal. Appl.* **397**, 255–259.

[12] Guo, X. and Zhang, Y. (2020). On risk-sensitive piecewise deterministic Markov decision processes. *Appl. Math. Optim.* **81**, 685–710.

[13] Guo, X.P. and Zhang, Y. (2017). Constrained total undiscounted continuous-time Markov decision processes. *Bernoulli* **23**, 1694–1736.

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

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

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

[17] Piunovskiy, A. (2015). Randomized and relaxed strategies in continuous-time Markov decision processes. *SIAM J. Control Optim.* **53**, 3503–3533.

[18] Serfozo, R. (1979). An equivalence between continuous and discrete time Markov decision processes. *Oper. Res.* **27**, 616–620.

[19] Sonderman, D. (1980). Comparing semi-Markov processes. *Math. Oper. Res.* **5**, 110–119.

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

[21] Zhang, Y. (2017). Continuous-time Markov decision processes with exponential utility. *SIAM J. Control Optim.* **55**, 2636-2660.