

A useful technique for piecewise deterministic Markov decision processes

Xin Guo ^{*} and Yi Zhang [†]

Abstract: This paper presents with justifications 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. As to be discussed and clarified, the auxiliary PDMDP possesses certain desired properties, which may not be possessed by the original PDMDP. Moreover, the performance measure of any policy in the original PDMDP can be replicated by the auxiliary PDMDP for a large class of performance criteria. As an application, we apply this technique to risk-sensitive PDMDPs with total cost criteria.

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

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

1 Introduction

This paper 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. Henceforth it will be termed as a piecewise deterministic Markov decision process (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, see e.g., [1, 3, 4, 7, 14, 15]. Consequently, 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 well appreciated that the theory of DTMDPs is better established when the underlying DTMDP model satisfies some compactness-continuity conditions, see [8, 9, 10]. One example of such compactness-continuity conditions is that the action space is a compact Borel space, the loss function 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. (A precise formulation of this is in the appendix.)

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

^{*}Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, U.K.. 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.

satisfy the desired continuity condition. We demonstrate this in Example A.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 A.1.

In the more specific context of continuous-time Markov decision processes (CTMDPs), where the state does not change between two consecutive jumps, another difficulty associated with nonstationary policies and transition intensities not strongly positive, which is different from the aforementioned one, was documented in [5, 6, 11].

The contribution of this paper is that we present a technique, which produces an auxiliary PDMDP model satisfying the following: a) the original PDMDP can be thinned out of the auxiliary PDMDP, and the performance of any policy in the original PDMDP is 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 conditions¹ 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 [7] under the extra requirement on the transition intensities being strongly positive, see the footnote in Proposition A.1. This requirement is omitted here with the help of the proposed technique.

In the context of CTMDPs, there have been known techniques, which produce new CTMDP models, in which the performance measure in the original model can be replicated. One example is the uniformization in [12, 13] requiring bounded transition intensities and justified under stationary policies. Its more recent variants that deal with models with unbounded transition intensities can be found in [2, 11], which are also similar in nature: the former one considers CTMDPs with a denumerable state space and deterministic stationary policies, whereas in the latter one, models with more general policies were considered. Our technique can be viewed as their extension from CTMDPs to PDMDPs, but there are some notable differences.

In greater detail, our technique is closer to [11] as we also consider general policies and unbounded transition intensities. Nevertheless, apart from a more general process under control being dealt with here, let us mention another more important difference from [11] as follows. The justification in [11] only serves the risk-neutral problem with total cost criteria, as it is based on the comparison of (total) occupation measures in the original model and the new model. This is not suitable for risk-sensitive problems, where in general the performance measures cannot be readily written as integrals of the cost rate with respect to the occupation measures. In contrast, here our justification applies to both risk-neutral and risk-sensitive problems with both total cost and average cost criteria, see Remark 3.1(c).

The rest of this paper is organized as follows. In Section 2 we describe the PDMDP model. In Section 2 we present and prove the main result with an application to risk-sensitive PDMDPs. The appendix presents the induced DTMDP. While the reference to the appendix can be avoided in the main text, it clarifies and demonstrates the issues mentioned in the beginning of this introduction.

2 Description of PDMDP model

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

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

for all $\Gamma_S \in \mathcal{B}(S)$, and

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

¹See Proposition A.1 for the precise definition of the compactness-continuity condition for the induced DTMDP.

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 $x \in S$

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

and $t \rightarrow \phi(x, t)$ is continuous. Unless stated otherwise, 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 $\mathcal{B}(A)$, endowed with the standard weak topology. For each $\mu \in \mathbb{P}(A)$,

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

Condition 2.1 For each $x \in S$, $\int_0^t \bar{q}_{\phi(x, s)} ds < \infty$, 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 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

$$h_n := (x_0, \theta_1, \dots, \theta_n, x_n); \quad t_n := t_{n-1} + \theta_n; \quad t_\infty(\omega) := \lim_{n \rightarrow \infty} t_n.$$

Then, (Ω, \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:

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

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 $\omega = (x_0, \theta_1, x_1, \theta_2, \dots) \in \Omega$, $t > 0$,

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

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^π 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

$$\begin{aligned} & P_x^\pi(\theta_{n+1} \in \Gamma_1, x_{n+1} \in \Gamma_2 | h_n) \\ &= \int_{\Gamma_1} e^{-\int_0^t \int_A q_{\phi(x_n, s)}(a)\pi_n(da|h_n, s) ds} \int_A \tilde{q}(\Gamma_2 | \phi(x_n, t), a)\pi_n(da|h_n, t) dt, \\ & \quad \forall \Gamma_1 \in \mathcal{B}((0, \infty)), \Gamma_2 \in \mathcal{B}(S); \\ & P_x^\pi(\theta_{n+1} = \infty, x_{n+1} = x_\infty | h_n) = e^{-\int_0^\infty \int_A q_{\phi(x_n, s)}(a)\pi_n(da|h_n, s) ds} \end{aligned} \quad (4)$$

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 reads

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

Here $\int_{t_n}^{t_{n+1}}$ is understood as $\int_{(t_n, t_{n+1}] \cap \mathbb{R}}$, 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.

However, our technique is also applicable to PDMDPs with other performance measure. For instance, one may consider the expected long run average cost defined by

$$\begin{aligned} \bar{V}(x, \pi) & := \overline{\lim}_{T \rightarrow \infty} E_x^\pi \left[\frac{\int_0^T \int_A c(\xi_t, a) \pi(da|\omega, t) dt}{T} \right] \\ & = \overline{\lim}_{T \rightarrow \infty} E_x^\pi \left[\frac{\sum_{n=0}^\infty \int_{t_n \wedge T}^{t_{n+1} \wedge T} \int_A c(\phi(x_n, s), a) \pi_n(da|h_n, s-t_n) ds}{T} \right], \end{aligned} \quad (5)$$

where $t_n \wedge T := \min\{t_n, T\}$. See Remark 3.1(c).

3 Main result

Fix $\lambda > 0$ in what follows. We introduce an auxiliary model $\{\check{S}, A, \check{q}, \check{\phi}, \check{c}\}$ defined in terms of the system primitives of the original model. 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.

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 $\check{S} = S \times \{-1, 1\}$, endowed with the product topology, where $\{-1, 1\}$ is with the discrete topology. The action space is A . The transition rate \check{q} on $\mathcal{B}(\check{S})$ given $\check{S} \times A$ is defined as follows:

$$\check{q}(dy \times \{-i\} | (x, i), a) = \lambda \delta_x(dy); \quad \check{q}(dy \times \{i\} | (x, i), a) = q(dy|x, a) - \lambda \delta_x(dy),$$

with $\delta_x(dy)$ being the Dirac measure concentrated on the singleton $\{x\}$, so that

$$\begin{aligned} \check{q}_{(x, i)}(a) & = \check{q}(\check{S} \setminus \{(x, i)\} | (x, i), a) = \check{q}(S \setminus \{x\} \times \{-1, 1\} | (x, i), a) + \check{q}(S \times \{-i\} | (x, i), a) \\ & = q_x(a) + \lambda \forall (x, i) \in \check{S}, a \in A. \end{aligned}$$

In other words, the auxiliary model has strongly positive transition intensities. The flow is defined by

$$\check{\phi}((x, i), t) = (\phi(x, t), i).$$

The cost rate is

$$\check{c}((x, i), a) = c(x, a) \forall (x, i) \in \check{S}, a \in A.$$

Let

$$\check{V}((x, i), \check{\pi}) = \check{E}_{(x, i)}^{\check{\pi}}[e^{\int_0^\infty \check{c}(\xi_t, a) \check{\pi}(da | \check{\omega}, t) dt}].$$

Definition 3.1 Consider the canonical sample space of the marked point process (\check{t}_n, x_n, i_n) , and a sample path

$$\check{\omega} = ((x_0, i_0), \check{\theta}_1, (x_1, i_1), \check{\theta}_2, \dots, (x_{n-1}, i_{n-1}), \check{\theta}_n, (x_n, i_n), \dots).$$

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, $x_l = \phi(x_{l-1}, \check{\theta}_l)$, where $\check{\theta}_l$ 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 (\check{t}_n, x_n, i_n) another marked point process $(\tau_{(m)}, x_{(m)}, i_{(m)})$ with $\tau_{(0)} := 0$ by counting only the points with marks immediately after honest jumps. 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 $(\tau_{(m)}, x_{(m)}, i_{(m)})$.

Theorem 3.1 Suppose Condition 2.1 is satisfied. For each policy $\pi = (\pi_n)$ in the original PDMDP model, there is a policy $\check{\pi} = (\check{\pi}_n)$ 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 $\check{P}_{(i, x)}^{\check{\pi}}$ coincides with the distribution of the marked point process (t_m, x_m) under P_x^π . In other words, the marked point process in the original model (under P_x^π) may be thinned out of (\check{t}_n, x_n, i_n) in the auxiliary model (under $\check{P}_{(x, i_0)}^{\check{\pi}}$) by counting only the points with marks immediately after honest jumps.

(b) $V(x, \pi) = \check{V}((x, i), \check{\pi})$.

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 $\check{\pi} = (\check{\pi}_n)$ in the auxiliary model defined as follows. For the n -history

$$\check{h}_n = ((x_0, i_0), \check{\theta}_1, (x_1, i_1), \check{\theta}_2, \dots, (x_{n-1}, i_{n-1}), \check{\theta}_n, (x_n, i_n))$$

in the auxiliary model, let $m = m(\check{h}_n)$ be the number of honest jumps over $(0, t_n]$ within \check{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 \check{h}_n . Then we define

$$\check{\pi}_n(da | \check{h}_n, t) = \pi_m(da | x_0, \tau_{(1)}, x_{(1)}, \tau_{(2)} - \tau_{(1)}, \dots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}, t + \check{t}_n - \tau_{(m)}) \quad (6)$$

for $t > 0$.

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

$$\begin{aligned} \check{\pi}(da | \check{\omega}, t) &= \check{\pi}_n(da | \check{h}_n, t - \check{t}_n) \\ &= \pi_m(da | x_0, \tau_{(1)}, x_{(1)}, \tau_{(2)} - \tau_{(1)}, \dots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}, t - \tau_{(m)}), \end{aligned} \quad (7)$$

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

For brevity, below we put

$$\begin{aligned}
\tilde{q}(dy|\phi(x_{(m)}, t), \pi_m) &:= \int_A \tilde{q}(dy|\phi(x_{(m)}, t), a)\pi_m(da|x_{(0)}, \tau_{(1)}, x_{(1)}, \tau_{(2)} - \tau_{(1)}, \dots, x_{(m)}, t); \\
q(dy|\phi(x_{(m)}, t), \pi_m) &:= \int_A q(dy|\phi(x_{(m)}, t), a)\pi_m(da|x_{(0)}, \tau_{(1)}, x_{(1)}, \tau_{(2)} - \tau_{(1)}, \dots, x_{(m)}, t); \\
q_{\phi(x_{(m)}, t)}(\pi_m) &:= \tilde{q}(S|\phi(x_{(m)}, t), \pi_m).
\end{aligned} \tag{8}$$

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

$$\begin{aligned}
&\check{P}_{(x, i)}^{\check{\pi}}(x_{(m+1)} \in \Gamma, \tau_{(m+1)} - \tau_{(m)} \in [0, T] | x_{(0)}, \tau_{(1)}, x_{(1)}, \dots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}) \\
&= \int_0^T \tilde{q}(\Gamma|\phi(x_{(m)}, t), \pi_m) e^{-\int_0^t q_{\phi(x_{(m)}, s)}(\pi_m) ds} dt
\end{aligned} \tag{9}$$

on $\{\tau_{(m)} < \infty\}$ for each $T > 0$, $\Gamma \in \mathcal{B}(S)$ and $m \geq 0$. Note that

$$\begin{aligned}
&\check{P}_{(x, i)}^{\check{\pi}}(x_{(m+1)} \in \Gamma, \tau_{(m+1)} - \tau_{(m)} \in [0, T] | x_{(0)}, \tau_{(1)}, x_{(1)}, \dots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}) \\
&= \sum_{n=0}^{\infty} \check{P}_{(x, i)}^{\check{\pi}}(x_{(m+1)} \in \Gamma, \tau_{(m+1)} - \tau_{(m)} \in [0, T], \\
&\quad \text{exactly } n \text{ fictitious jumps over } [\tau_{(m)}, \tau_{(m+1)}] | x_{(0)}, \tau_{(1)}, x_{(1)}, \dots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}).
\end{aligned}$$

Since $\sum_{n \geq 0} \frac{\lambda^n t^n}{n!} = e^{\lambda t}$, equality (9) would be justified once we show that

$$\begin{aligned}
&\check{P}_{(x, i)}^{\check{\pi}}(x_{(m+1)} \in \Gamma, \tau_{(m+1)} - \tau_{(m)} \in [0, T], \\
&\quad \text{exactly } n \text{ fictitious jumps over } [\tau_{(m)}, \tau_{(m+1)}] | x_{(0)}, \tau_{(1)}, x_{(1)}, \dots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}) \\
&= \int_0^T \frac{\lambda^n t^n}{n!} \tilde{q}(\Gamma|\phi(x_{(m)}, t), \pi_m) e^{-\lambda t} e^{-\int_0^t q_{\phi(x_{(m)}, s)}(\pi_m) ds} dt.
\end{aligned} \tag{10}$$

For this, let us verify for each $T > 0$ and $y \in S$ that

$$\begin{aligned}
f(T, y, n-1) &:= \int_0^T \int_0^{T-r_1} \int_0^{T-\sum_{i=1}^2 r_i} \dots \int_0^{T-\sum_{i=1}^{n-1} r_i} \tilde{q}(\Gamma|\phi(y, \sum_{i=1}^{n-1} r_i + t), \pi_m) \\
&\quad \times e^{-\int_0^t (q_{\phi(y, \sum_{i=1}^{n-1} r_i + s)}(\pi_m) + \lambda) ds} \lambda^{n-1} e^{-\int_0^{\sum_{i=1}^{n-1} r_i} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} dt dr_{n-1} \dots dr_1 \\
&= \int_0^T \frac{\lambda^{n-1} t^{n-1}}{(n-1)!} \tilde{q}(\Gamma|\phi(y, t), \pi_m) e^{-\int_0^t (q_{\phi(y, s)}(\pi_m) + \lambda) ds} dt
\end{aligned} \tag{11}$$

for each $n \geq 1$, where the notation in (8) is in use with $x_{(m)}$ being replaced by y . This would yield the desired relation (10) because by (7)

$$\begin{aligned}
f(T, x_{(m)}, n-1) &= \check{P}_{(x, i)}^{\check{\pi}}(x_{(m+1)} \in \Gamma, \tau_{(m+1)} - \tau_{(m)} \in [0, T], \\
&\quad \text{exactly } n-1 \text{ fictitious jumps over } [\tau_{(m)}, \tau_{(m+1)}] | x_{(0)}, \tau_{(1)}, x_{(1)}, \dots, \tau_{(m)} - \tau_{(m-1)}, x_{(m)}).
\end{aligned}$$

Relation (11) holds trivially when $n = 1$ because

$$f(T, y, 0) = \int_0^T \tilde{q}(\Gamma|\phi(y, t), \pi_m) e^{-\int_0^t (q_{\phi(y, s)}(\pi_m) + \lambda) ds} dt$$

by definition. Suppose (11) holds. Then

$$\begin{aligned}
f(T, y, n) &= \int_0^T \lambda \int_0^{T-r_1} \int_0^{T-\sum_{i=1}^2 r_i} \cdots \int_0^{T-\sum_{i=1}^{n-1} r_i} \int_0^{T-\sum_{i=1}^n r_i} \tilde{q}(\Gamma|\phi(y, \sum_{i=1}^n r_i + t), \pi_m) \\
&\times e^{-\int_0^t (q_{\phi(y, \sum_{i=1}^n r_i + s)}(\pi_m) + \lambda) ds} \lambda^{n-1} e^{-\int_0^{\sum_{i=1}^n r_i} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} dt dr_n \dots dr_2 dr_1 \\
&= \int_0^T \lambda e^{-\int_0^{r_1} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} \\
&\times \left\{ \int_0^{T-r_1} \int_0^{T-r_1-r_2} \cdots \int_0^{T-r_1-\sum_{i=2}^n r_i} \tilde{q}(\Gamma|\phi(\phi(y, r_1), \sum_{i=2}^n r_i + t), \pi_m) \right. \\
&\times \left. e^{-\int_0^t (q_{\phi(\phi(y, r_1), \sum_{i=2}^n r_i + s)}(\pi_m) + \lambda) ds} \lambda^{n-1} e^{-\int_0^{\sum_{i=2}^n r_i} (q_{\phi(\phi(y, r_1), s)}(\pi_m) + \lambda) ds} dt dr_n \dots dr_2 \right\} dr_1
\end{aligned}$$

where the second equality holds because of (1) and that

$$\begin{aligned}
e^{-\int_0^{\sum_{i=1}^n r_i} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} &= e^{-\int_0^{r_1} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} e^{-\int_{r_1}^{\sum_{i=1}^n r_i} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} \\
&= e^{-\int_0^{r_1} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} e^{-\int_0^{\sum_{i=2}^n r_i} (q_{\phi(\phi(y, r_1), s)}(\pi_m) + \lambda) ds}.
\end{aligned}$$

Now after applying the induction supposition (11) (with y being replaced by $\phi(y, r_1)$, T being replaced by $T - r_1$, and $\sum_{i=2}^n r_i$ playing the same role as $\sum_{i=1}^{n-1} r_i$ therein) to the previous inner integral, we see that

$$\begin{aligned}
f(T, y, n) &= \int_0^T \lambda e^{-\int_0^{r_1} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} \\
&\times \left\{ \int_0^{T-r_1} \frac{\lambda^{n-1} t^{n-1}}{(n-1)!} \tilde{q}(\Gamma|\phi(\phi(y, r_1), t), \pi_m) e^{-\int_0^t (q_{\phi(\phi(y, r_1), s)}(\pi_m) + \lambda) ds} dt \right\} dr_1 \\
&= \int_0^T \left\{ \int_0^{T-r_1} \frac{\lambda^n t^{n-1}}{(n-1)!} \tilde{q}(\Gamma|\phi(y, r_1 + t), \pi_m) e^{-\int_0^{t+r_1} (q_{\phi(x, s)}(\pi_m) + \lambda) ds} dt \right\} dr_1 \\
&= \int_0^T \int_0^t \frac{\lambda^n (t-r_1)^{n-1}}{(n-1)!} \tilde{q}(\Gamma|\phi(y, t), \pi_m) e^{-\int_0^t (q_{\phi(x, s)}(\pi_m) + \lambda) ds} dr_1 dt \\
&= \int_0^T \frac{\lambda^n t^n}{n!} \tilde{q}(\Gamma|\phi(y, t), \pi_m) e^{-\int_0^t (q_{\phi(x, s)}(\pi_m) + \lambda) ds} dt,
\end{aligned}$$

as desired, where the second equality follows from (1) and that

$$e^{-\int_0^{r_1} (q_{\phi(y, s)}(\pi_m) + \lambda) ds} e^{-\int_0^t (q_{\phi(\phi(y, r_1), s)}(\pi_m) + \lambda) ds} = e^{-\int_0^{t+r_1} (q_{\phi(y, s)}(\pi_m) + \lambda) ds},$$

and the third equality is by a change of variable and interchanging legitimately the order of integration. Part (a) is thus proved.

(b) It follows from part (a) that

$$\begin{aligned}
\check{V}((x, i_0), \check{\pi}) &= \check{E}_{(x, i_0)}^{\check{\pi}} [e^{\int_0^\infty \check{c}(\xi_t, a) \check{\pi}(da|\check{\omega}, t) dt}] \\
&= \check{E}_{(x, i_0)}^{\check{\pi}} \left[e^{\sum_{m=0}^\infty \int_{\tau(m)}^{\tau(m+1)} \int_A c(\phi(x_m, s), a) \pi_m(da|x, \tau(1), x(2), \tau(2) - \tau(1), \dots, \tau(m) - \tau(m-1), x_m, s - \tau(m)) ds} \right] \\
&= E_{(x, i_0)}^\pi \left[e^{\sum_{m=0}^\infty \int_{t_m}^{t_m+1} \int_A c(\phi(x_m, s), a) \pi_m(da|x, t_1, x_2, t_2 - t_1, \dots, t_m - t_{m-1}, x_m, s - t_m) ds} \right] = V(x, \pi),
\end{aligned}$$

where the second equality holds by (7) and the definition of \check{c} . The proof is completed. \square

Remark 3.1 (a) By Theorem 3.1(b), $\check{V}^*(x) \leq V^*(x)$ for each $x \in S$.

(b) By inspecting the proof of Theorem 3.1 (see especially (6) and (7) 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 , $\check{V}((x, i), \varphi) = V(x, \varphi)$ 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 $V^*(x) = \check{V}^*(x) = \check{V}((x, i), \varphi) = V(x, \varphi)$ for each $x \in S$.

(c) By Theorem 3.1(a) and the second equality in (5), we see $\bar{V}(x, \pi) = \check{\bar{V}}((x, i), \check{\pi})$ for the same policies π and $\check{\pi}$ in Theorem 3.1, too.

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 Borel space.

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

Lemma 3.1 If the original PDMDP model satisfies Conditions 3.1 and 2.1, then the auxiliary model satisfies the corresponding versions of Conditions 3.1 and 2.1, too.

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

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

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

The following statement was obtained in Theorem 3.1 and Remark 3.1 of [7].

Proposition 3.1 Suppose Conditions 2.1 and 3.1 are satisfied. In addition², $\inf_{(x, a) \in S \times A} q_x(a) > 0$. Then the following assertions hold.

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

$$\begin{aligned} & -(V(\phi(x, t)) - V(x)) \tag{12} \\ &= \int_0^t \inf_{a \in A} \left\{ \int_S V(y) \tilde{q}(dy|\phi(x, \tau), a) - (q_{\phi(x, \tau)}(a) - c(\phi(x, \tau), a)) V(\phi(x, \tau)) \right\} d\tau \\ & \forall t \in [0, \infty), x \in S^* := \{x \in S : V^*(x) < \infty\}; V(x) < \infty \forall x \in S^*; V(x) = \infty \forall x \notin S^*. \end{aligned}$$

In particular, $V^*(\phi(x, t))$ is absolutely continuous in t for each $x \in S^*$.

²See the footnote in Proposition A.1.

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

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

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

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 requiring $\inf_{(x,a) \in S \times A} q_x(a) > 0$.*

Proof. The statement follows from Remark 3.1 and applying Proposition 3.1 to the auxiliary model, 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 $\check{V}^*((x, i))$ in the auxiliary PDMDP model depends on (x, i) only through x , and can thus be identified as $\check{V}^*(x)$.

For this, we will apply the following result from [7]: 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

$$\begin{aligned} \mathcal{T} \circ V(x) &:= \inf_{\rho \in \mathcal{R}} \left\{ \int_0^\infty e^{-\int_0^\tau (q_{\phi(x,s)}(\rho_s) - c(\phi(x,s), \rho_s)) ds} \left(\int_S V(y) \tilde{q}(dy|\phi(x, \tau), \rho_\tau) \right) d\tau \right. \\ &\quad \left. + e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s) ds} e^{\int_0^\infty c(\phi(x,s), \rho_s) ds} \right\}, \quad \forall x \in S, \end{aligned} \quad (13)$$

Here and below, \mathcal{R} is the space of $\mathbb{P}(A)$ -valued measurable mappings $\rho = (\rho_t(da))$ on $(0, \infty)^3$, and $e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s) ds} e^{\int_0^\infty c(\phi(x,s), \rho_s) ds} := 0$ whenever $e^{-\int_0^\infty q_{\phi(x,s)}(\rho_s) ds} = 0$.

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

According to Lemma 3.1 and that the transition intensities in the auxiliary model are strongly positive (as $\inf_{x \in S, a \in A} \check{q}_{(x,i)}(a) \geq \lambda > 0$), we may apply the result just quoted above to the auxiliary model and conclude that \check{V}^* is the minimal $[1, \infty]$ -valued measurable function to the following equation

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

³Two elements in \mathcal{R} that coincide almost everywhere are not distinguished.

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

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

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

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

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

does not involve $i \in \{-1, 1\}$, where the equality holds by the definition of \tilde{q} and \check{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 $\check{V}^*(x) = V^*(x)$. Note that the optimality equations for both the original model and the auxiliary model are the same and given by (12), the statement of this corollary follows from applying again Proposition 3.1 to the auxiliary model. \square

A Appendix: Induced DTMDP

We shall formulate a DTMDP model induced from the PDMDP model with total cost criteria 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. This is to serve the formulation of Example A.1 below.

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} was defined in the proof of Corollary 3.1. We endow \mathcal{R} with the Young topology.⁴

⁴The Young topology on \mathcal{R} is the weakest topology with respect to which the function $\rho \in \mathcal{R} \rightarrow \int_0^\infty \int_A f(t, a) \rho_t(da) dt$ is continuous 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., $\int_0^\infty \sup_{a \in A} |f(t, a)| dt < \infty$.

- 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

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

(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 by

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

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.

Proposition A.1 *Under Conditions 2.1 and 3.1, for each $(\theta, x) \in \mathbf{X}$ and $(\tau, y) \in \mathbf{X}$, $\rho \in \mathbf{A} \rightarrow l((\theta, x), \rho, (\tau, y))$ is lower semicontinuous, and \mathbf{A} is a compact Borel space. If in addition, the transition intensities are strongly positive⁵, then for each $(\theta, x) \in \mathbf{X}$, the function $\rho \in \mathbf{A} \rightarrow \int_{\mathbf{X}} f(z) p(dz | (\theta, x), \rho)$ is continuous for each bounded measurable function f on \mathbf{X} .*

The next example shows that if the transition intensities are not strongly positive, then it can happen that $\rho \in \mathbf{A} \rightarrow \int_{\mathbf{X}} f(z) p(dz | (\theta, x), \rho)$ is not continuous for some bounded measurable function f on \mathbf{X} .

Example A.1 *Suppose S is any finite set (endowed with discrete topology), and $A = [0, 1]$, which is a compact Borel 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 $\rho \in \mathbf{A}$ and $(\rho^{(n)}) \subseteq \mathbf{A}$ 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 $g(t, a)$,*

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

as $n \rightarrow \infty$, by using the dominated convergence theorem. Thus, $\rho^{(n)} \rightarrow \rho$ as $n \rightarrow \infty$. (Recall that \mathbf{A} is endowed with the Young topology.) Now for $f(t, x) \equiv 0$ on $(0, \infty) \times S$ and $f(\infty, x_\infty) = 1$,

$$\int_{\mathbf{X}} f(z) p(dz | (\theta, x), \rho^{(n)}) = e^{-\int_0^\infty q_x(\rho_s^{(n)}) ds} = e^{-\int_0^\infty \frac{1}{n} ds} = 0 < 1 = e^{-\int_0^\infty 0 ds} = \int_{\mathbf{X}} f(z) p(dz | (\theta, x), \rho).$$

References

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

⁵This requirement was unfortunately missing and overlooked in [7]. Indeed, the proof of Lemma 4.1 of [7] 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 A.1. Since the rest of the arguments in [7] are largely based on that lemma, this missing requirement on the strong positivity of transition intensities should be added in [7] wherever appropriate. We thank Dr Yonghui Huang (Sun Yat-Sen University, China) for drawing our attention on this inaccuracy.

- [2] Çekyay, B. (2018). Customizing exponential semi-Markov decision processes under the discounted cost criterion. *Eur. J. Oper. Res.* **266**, 168–178.
- [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. (2004). Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Oper. Res.* **29**, 492–524.
- [6] Guo, X.P. and Zhang, Y. (2017). Constrained total undiscounted continuous-time Markov decision processes. *Bernoulli* **23**, 1694–1736.
- [7] Guo, X. and Zhang, Y. (2020). On risk-sensitive piecewise deterministic Markov decision processes *Appl. Math. Optim.* **81**, 685–710.
- [8] Hernández-Lerma, O. and Lasserre, J. (1996). *Discrete-Time Markov Control Processes*. Springer-Verlag, New York.
- [9] Hernández-Lerma, O. and Lasserre, J. (1999). *Further Topics in Discrete-Time Markov Control Processes*, Springer-Verlag, New York.
- [10] Jaśkiewicz, A. (2008). A note on negative dynamic programming for risk-sensitive control. *Oper. Res. Lett.* **36**, 531–534.
- [11] Piunovskiy, A. (2015). Randomized and relaxed strategies in continuous-time Markov decision processes. *SIAM J. Control Optim.* **53**, 3503–3533.
- [12] Puterman, M. (1994). *Markov Decision Processes*. Wiley, New York.
- [13] Serfozo, R. (1979). An equivalence between continuous and discrete time Markov decision processes. *Oper. Res.* **27**, 616–620.
- [14] Yushkevich, A.A. (1980). On reducing a jump controllable Markov model to a model with discrete time. *Theory Probab. Appl.* **25**, 58–68.
- [15] Zhang, Y. (2017). Continuous-time Markov decision processes with exponential utility. *SIAM J. Control Optim.* **55**, 2636–2660.