On gradual-impulse control of continuous-time Markov decision processes with exponential utility

Xin Guo[*], Aiko Kurushima[†], Alexey Piunovskiy[‡] and Yi Zhang [§]

Abstract: We consider a 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 show, under natural conditions on the system primitives, the existence of a deterministic stationary optimal policy out of a more general class of policies that allow multiple simultaneous impulses, randomized selection of impulses with random effects, and accumulation of jumps. After characterizing the value function using the optimality equation, we reduce the 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.

Keywords: Continuous-time Markov decision processes. Exponential utility. Impulse-gradual control. Risk-sensitive criterion. Optimality equation.

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

1 Introduction

This paper considers a gradual-impulse control problem for continuous-time Markov decision processes (CTMDPs) with the performance to be minimized being the expected exponential utility of the total cost. In this model, the decision maker can control the process gradually via its local characteristics (transition rate), and also has the option of affecting impulsively the state of the process. The system dynamics is depicted in Figure 1 below.

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 the present paper.

Example 1.1 *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.*

[*] Department of Mathematical Sciences, University of Liverpool, Liverpool, U.K.. E-mail: x.guo21@liv.ac.uk.

[†] Department of Economics, Sophia University, Tokyo, Japan. Email: kurushima@sophia.ac.jp.

[‡] Department of Mathematical Sciences, University of Liverpool, Liverpool, U.K.. E-mail: piunov@liv.ac.uk.

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

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 3.1 below, which demonstrates the situations when applying only one impulse is insufficient for optimality.

Most previous works on gradual-impulse control do not allow multiple simultaneous impulses at a single time moment, see [3, 4, 6, 14, 17, 18, 21]. Extra conditions are imposed therein to guarantee there is no need to apply more than one impulse at a single time moment. Example 1.1 described a situation when those conditions are not satisfied. There is convenience if only one impulse is allowed at a given time moment, because at each time moment, there is only one state, so that one can construct the process under control in the (original) state space of the gradual-impulse control problem. When there is only gradual control, it is convenient to construct the CTMDP using a marked point process with the mark space being the same as the state space of the original control problem. If multiple impulses were applied in a sequence at a single time moment, then there would be multiple states associated with the single time moment. If one wishes to construct the problem using a marked point process, then the mark space must be enlarged, so that a sequence of impulses applied at the single time moment and the post-impulse states are merged as a single “mark”, which will be called an intervention. Necessarily this leads to a more complicated marked point process with each mark corresponding to a sample path of a discrete-time Markov decision process (DTMDP). This idea was employed and implemented in [7].

Another way of constructing rigorously a gradual-impulse control problem of CTMDPs admitting multiple simultaneous impulses comes from [24]. The idea is to keep the original state space, but to enlarge the time $t \in [0, \infty)$ to (n, t) with the first coordinate, roughly speaking, counting the number of impulses applied at the time t. Consequently, several concepts about stochastic processes needed be extended.

In the present work, we follow the construction of [7] but with more general control policies. Compared to the previous literature on impulse or gradual-impulse control problems of CTMDPs, to the best of our knowledge, we consider the most general setup: the policy allows to make relaxed gradual controls and randomized impulsive controls with randomized consequences, multiple simultaneous impulses are allowed, and accumulation of jumps of the process is not excluded. We study the gradual-impulse control problem of CTMDPs with the system performance measure being the expectation of the exponential utility of the total cost to be minimized. For risk-sensitive CTMDPs with gradual control only and total or average cost criteria, see e.g., [11, 12, 16, 19, 22, 25]. In close relation to the present paper, the risk-sensitive optimal stopping problem of a continuous-time Markov chain was recently considered in [1], which is a special impulse control problem but with a more general utility function.

The main optimality results of this paper lie in the following. We characterize the value function of the gradual-impulse control problem for CTMDPs in terms of the optimality equation, and show the existence of deterministic stationary optimal policies, under quite general and natural conditions compared to the literature. For example, the growth on the gradual cost rates and impulse cost functions, as well as the transition rate can be quite general. In comparison, only bounded transition and cost rates were allowed in [7], which deals with a discounted problem with linear utility. The boundedness conditions were to guarantee that the Dynkin formula is applicable to functions of interest therein, which is important for the argument therein.

The method of investigations in the present paper is different from [7], but is closer to [25], which studies a similar problem for CTMDPs but with gradual control only. Although both the present paper and [25] follow the same idea of reducing the original problem to a DTMDP, the implementation for the gradual-impulse control problem becomes more involving. In particular, the connection between a strategy in the induced DTMDP and a policy in the gradual-impulse control problem, which is at the core of the justification of the reduction, becomes more delicate, see Subsection 4.2 below.

In the induced DTMDP, an action is a triplet, including the time until the next time of applying an impulse (if no natural jump occurred by then), the next impulse itself, and the decision rule for the selection of gradual controls. Apart from being with a more complicated action space than the original problem, the induced DTMDP model is not so convenient. For example, it is not a semicontinuous model even if the system primitives of the gradual-impulse control problem satisfy the compactness-continuity conditions. Consequently, the existence of an optimal policy does not follow automatically from the reduction to this DTMDP. In this connection, we mention that Lemma 5.12 of [25] is inaccurate unless further conditions were imposed therein. Here we incidentally show that the optimality results in [25] remain correct in spite of that error. Accordingly, the second step in the investigation is to reduce further the DTMDP model to yet another one, which is a semicontinuous model, and with a simple action space (the union of the set of gradual actions and impulses). This second reduction is done based on the investigation of the optimality equation of the DTMDP obtained from the first reduction.

The rest of the paper is organized as follows. We present the rigorous construction of the controlled process and problem statement in Section 2. Section 3 consists of the main optimality results, whose proof is postponed to Section 5. The argument is based on the connection with a DTMDP model, which is introduced in Section 4. The paper is finished with a conclusion in Section 6. To improve the readability, we summarize the relevant notions and facts about DTMDPs in the appendix.

Notation and conventions. In what follows, $\mathcal{B}(X)$ is the Borel σ-algebra of the topological space X, I stands for the indicator function, and $\delta_x(\cdot)$ is the Dirac measure concentrated on the singleton $\{x\}$, assumed to be measurable. A measure is σ-additive and $[0,\infty]$-valued. Here and below, unless stated otherwise, the term of measurability is always understood in the Borel sense. Throughout this paper, we adopt the conventions of $\frac{0}{0}:=0,\ 0\cdot\infty:=0,\ \frac{1}{0}:=+\infty,\ \infty-\infty:=\infty$. For each function f on X, let $||f||:=\sup_{x\in X}|f(x)|$.

2 Model description and problem statement

2.1 System primitives of the gradual-impulse control problem

We describe the primitives of the model as follows. The state space is $\mathbf{X}$, the space of gradual controls is $\mathbf{A}^G$, and the space of impulsive controls is $\mathbf{A}^I$. It is assumed that $\mathbf{X}$, $\mathbf{A}^G$ and $\mathbf{A}^I$ are all Borel spaces, endowed with their Borel σ-algebras $\mathcal{B}(\mathbf{X})$, $\mathcal{B}(\mathbf{A}^G)$ and $\mathcal{B}(\mathbf{A}^I)$, respectively. The transition rate, on which the gradual control acts, is given by $q(dy|x,a)$, which is a signed kernel from $\mathbf{X}\times\mathbf{A}^G$, endowed with its Borel σ-algebra, to $\mathcal{B}(\mathbf{X})$, satisfying the following conditions: $q(\Gamma|x,a)\in[0,\infty)$ for each $\Gamma\in\mathcal{B}(\mathbf{X}), x\notin\Gamma$; $q(\mathbf{X}|x,a)=0,\ x\in\mathbf{X},\ a\in\mathbf{A}^G$; $\bar{q}_x:=\sup_{a\in\mathbf{A}^G}q_x(a)<\infty,\ x\in\mathbf{X}$, where $q_x(a):=-q(\{x\}|x,a)$ for each $(x,a)\in\mathbf{X}\times\mathbf{A}^G$. For notational convenience, we introduce $\tilde{q}(dy|x,a):=q(dy\setminus\{x\}|x,a),\ \forall\ x\in\mathbf{X},\ a\in\mathbf{A}^G$. If the current state is $x\in\mathbf{X}$, 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 $\mathbf{X}\times\mathbf{A}^I$ to $\mathcal{B}(\mathbf{X})$. Finally, given the current state $x\in\mathbf{X}$, 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 $\mathbf{X}\times\mathbf{A}^G$ and $\mathbf{X}\times\mathbf{A}^I\times\mathbf{X}$, respectively. Throughout this paper, we assume that $\mathbf{A}^G$ and $\mathbf{A}^I$ are compact Borel spaces. It is without loss of generality to regard $\mathbf{A}^G$ and $\mathbf{A}^I$ as two disjoint compact subsets of a Borel space $\tilde{\mathbf{A}}=\mathbf{A}^G\cup\mathbf{A}^I$. Furthermore, we assume that

$$\sup_{a\in\mathbf{A}^G}c^G(x,a)<\infty,\ \forall x\in\mathbf{X}. \tag{1}$$

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 $\mathbf{X}$, 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 $\mathbf{X}$. 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 $\mathbf{X}$ at a single time moment. Moreover, the order of the impulses and their resulting states are 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 [7], 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 subsection.

2.2 Definition and interpretation of an intervention

At the beginning of an intervention, the decision maker chooses whether to apply an impulse, and which one to apply. If the current state is $x \in \mathbf{X}$, after an impulse $b \in \mathbf{A}^I$ is chosen, the new state say $y \in \mathbf{X}$ 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 $\Delta \notin \mathbf{A}^I \cup \mathbf{X}$ is artificially fixed, which is chosen when the decision maker decides not to apply any more impulse at the current instant, 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 itself a sequential decision process. More precisely, an intervention can be regarded as a trajectory 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 2.1 *The intervention DTMDP model is specified by the tuple $\{\mathbf{X}_\Delta, \mathbf{A}^I_\Delta, Q\}$, which is defined in terms of the primitives of the gradual-impulse control problem given in Subsection 2.1, where the state space is $\mathbf{X}_\Delta := \mathbf{X} \cup \{\Delta\}$ with $\Delta \notin \boldsymbol{X} \cup \boldsymbol{A}^I$ being a cemetery point, the action space is $\mathbf{A}^I_\Delta := \mathbf{A}^I \cup \{\Delta\}$, the one-step transition probability from $\boldsymbol{X}_\Delta \times \boldsymbol{A}^I_\Delta$ to $\mathcal{B}(\boldsymbol{X}_\Delta)$ is $Q(dy|x,b)$. Here 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 $\mathbf{X}$. Then its canonical sample space is $\mathbf{Y} := (\bigcup_{k=0}^\infty \mathbf{Y}_k) \cup (\mathbf{X} \times \mathbf{A}^I)^\infty$, where for each $\infty > k \geq 1$, $\mathbf{Y}_k := (\mathbf{X} \times \mathbf{A}^I)^k \times (\mathbf{X} \times \{\Delta\}) \times (\{\Delta\} \times \{\Delta\})^\infty$, and $\mathbf{Y}_0 := (\mathbf{X} \times \{\Delta\}) \times (\{\Delta\} \times \{\Delta\})^\infty$. Here, if $y \in \mathbf{Y}_k$, $\infty > k \geq 0$, then there are k impulses applied in the intervention y. Similarly, if $y \in (\mathbf{X} \times \mathbf{A}^I)^\infty$, then there are infinitely many impulses applied in the intervention y.

Now we give the following definition.

Definition 2.2 *An intervention is an element of $\boldsymbol{Y}$.*

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 notation 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 $x \in \mathbf{X}$, if the controller decides to use Δ, then it means, no more impulse is used at this instant, and the intervention DTMDP is absorbed at Δ in the next step; 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 $\mathbf{Y}^* := \mathbf{Y} \setminus \mathbf{Y}_0 = (\bigcup_{k=1}^{\infty} \mathbf{Y}_k) \cup (\mathbf{X} \times \mathbf{A}^I)^{\infty}$ be the set of interventions, where some impulses are applied.

In an intervention, locally, the selection of impulses (including the "pseudo" impulse Δ) from $\mathbf{A}^I_\Delta$ 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 2.3 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. We refer the reader to the appendix for standard terminologies of DTMDPs. The way how a strategy in the intervention DTMDP model is incorporated into a policy in Definition 2.3 below is through its strategic measure. Let $\beta^\sigma(\cdot|x)$ denote the corresponding strategic measure of a strategy σ of the intervention DTMDP, given the initial state $x \in \mathbf{X}$. By the Ionescu-Tulcea theorem, see e.g., Proposition C.10 in [13], the mapping $x \in \mathbf{X} \to \beta^\sigma(\cdot|x)$ 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 $x \in \mathbf{X}$. Let $\mathcal{P}^{\mathbf{Y}^*} := \{\beta(\cdot|\cdot) \in \mathcal{P}^{\mathbf{Y}} : \ \beta(\mathbf{Y}^*|x) = 1, \ \forall\ x \in \mathbf{X}\}$, and for each $x \in \mathbf{X}$, $\mathcal{P}^{\mathbf{Y}^*}(x) := \{\beta(\cdot|x) : \beta(\cdot|\cdot) \in \mathcal{P}^{\mathbf{Y}}, \ \beta(\mathbf{Y}^*|x) = 1\}$.

2.3 Construction of the controlled process

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

Let $\mathbf{Y}_\Delta := \mathbf{Y} \cup \{\Delta\}$, $\Omega_0 := \mathbf{Y} \times (\{0\} \times \mathbf{Y}) \times (\{\infty\} \times \{\Delta\})^{\infty}$, and $\Omega_n := \mathbf{Y} \times (\{0\} \times \mathbf{Y}) \times ((0,\infty) \times \mathbf{Y})^n \times (\{\infty\} \times \{\Delta\})^{\infty}$ for all $n = 1, 2, \ldots$. The canonical space Ω is defined as $\Omega := (\bigcup_{n=0}^{\infty} \Omega_n) \cup \big(\mathbf{Y} \times ((0,\infty) \times \mathbf{Y})^{\infty}\big)$ 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: $\omega = (y_0, \theta_1, y_1, \theta_2, y_2, \ldots)$. 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

$$y_0 := (x_0, \Delta, \Delta, \ldots), \ \theta_1 \equiv 0. \tag{2}$$

The sequence $\{\theta_n\}_{n=1}^{\infty}$ 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. (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 (2). The collection of all such partial 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 by $T_n(\omega) := \sum_{i=1}^{n} \Theta_i(\omega) = \sum_{i=1}^{n} \theta_i$ and $T_\infty(\omega) := \lim_{n\to\infty} T_n(\omega)$ for all $\omega \in \Omega$. Let $H_n := (Y_0, \Theta_1, Y_1, \ldots, \Theta_n, Y_n)$. Finally, we define the controlled process $\{\xi_t\}_{t\in[0,\infty)}$ by

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

It is convenient to introduce the random measure μ of the marked point process $\{(T_n, Y_n)\}_{n=1}^{\infty}$ on $(0,\infty) \times \mathbf{Y}$: $\mu(dt \times dy) = \sum_{n\ge 2} I_{\{T_n<\infty\}} \delta_{(T_n,Y_n)}(dt \times dy)$. Let $\mathcal{F}_t := \sigma\{H_1\} \vee \sigma\{\mu((0,s] \times B) : s \le t, B \in \mathcal{B}(\mathbf{Y})\}$ for $t \in [0,\infty)$.

We will use the following notation in the next definition. For each intervention

$$y = (x_0, b_0, x_1, b_1, \ldots) \in \mathbf{Y},$$

define $\bar{x}(y) := x_k$ if $\infty > k = 0, 1, \ldots$ is the unique integer such that $y \in \mathbf{Y}_k$ (if $k \ge 1$, 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 $\bar{x}(y) := \Delta$. The 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 2.3 *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 $\mathbf{H}_n$ such that $\Phi_n(\{\infty\}|h_n) = 1$ if $y_n \in (\boldsymbol{X} \times \boldsymbol{A}^I)^\infty$, Π_n is a stochastic kernel on $\mathbf{A}^G$ given $\mathbf{H}_n \times (0,\infty)$, Γ_n^0 is a stochastic kernel on $\mathbf{Y}$ given $\mathbf{H}_n \times (0,\infty) \times \mathbf{X}$ satisfying $\Gamma_n^0(\cdot|h_n, t, x) \in \mathcal{P}^{\boldsymbol{Y}}(x)$ for each $h_n \in \boldsymbol{H}_n$, $x \in \boldsymbol{X}$ and $t \in (0,\infty)$, and Γ_n^1 is a stochastic kernel on $\mathbf{Y}$ given $\mathbf{H}_n$ satisfying $\Gamma_n^1(\cdot|h_n) \in \mathcal{P}^{\boldsymbol{Y}^*}(\bar{x}(y_n))$ for each $h_n \in \boldsymbol{H}_n$. (The above conditions apply when $y_n \ne \Delta$; otherwise, all the values of Φ_n, Π_n, Γ_n^0 and Γ_n^1 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 $\int_{\mathbf{A}^G} q_{\bar{x}(Y_n)}(a)\Pi_n(da|H_n, t)$. Below, we put $q_\Delta(a) := 0$ for each $a \in \mathbf{A}^G$. In other words, Π_n is the (decision rule of) relaxed gradual control. Given that 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 (3) and (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.

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 $\mathbb{P}_{x_0}^u$ on $(\Omega, \mathcal{F})$, where $x_0 \in \mathbf{X}$ is the fixed initial state of the system dynamics. For brevity, we introduce the following notations for each $n \ge 1$, $\Gamma \in \mathcal{B}(\mathbf{X})$ and $h_n = (y_0, \theta_1, y_1, \ldots, \theta_n, y_n) \in \mathbf{H}_n$:

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

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

$$G_n^u(\{+\infty\} \times \{\Delta\}|h_n) := \delta_{y_n}(\{\Delta\}) + \delta_{y_n}(\mathbf{Y})e^{-\Lambda_n^u(\mathbf{X}|h_n,+\infty)}\Phi_n(\{+\infty\}|h_n), \tag{3}$$

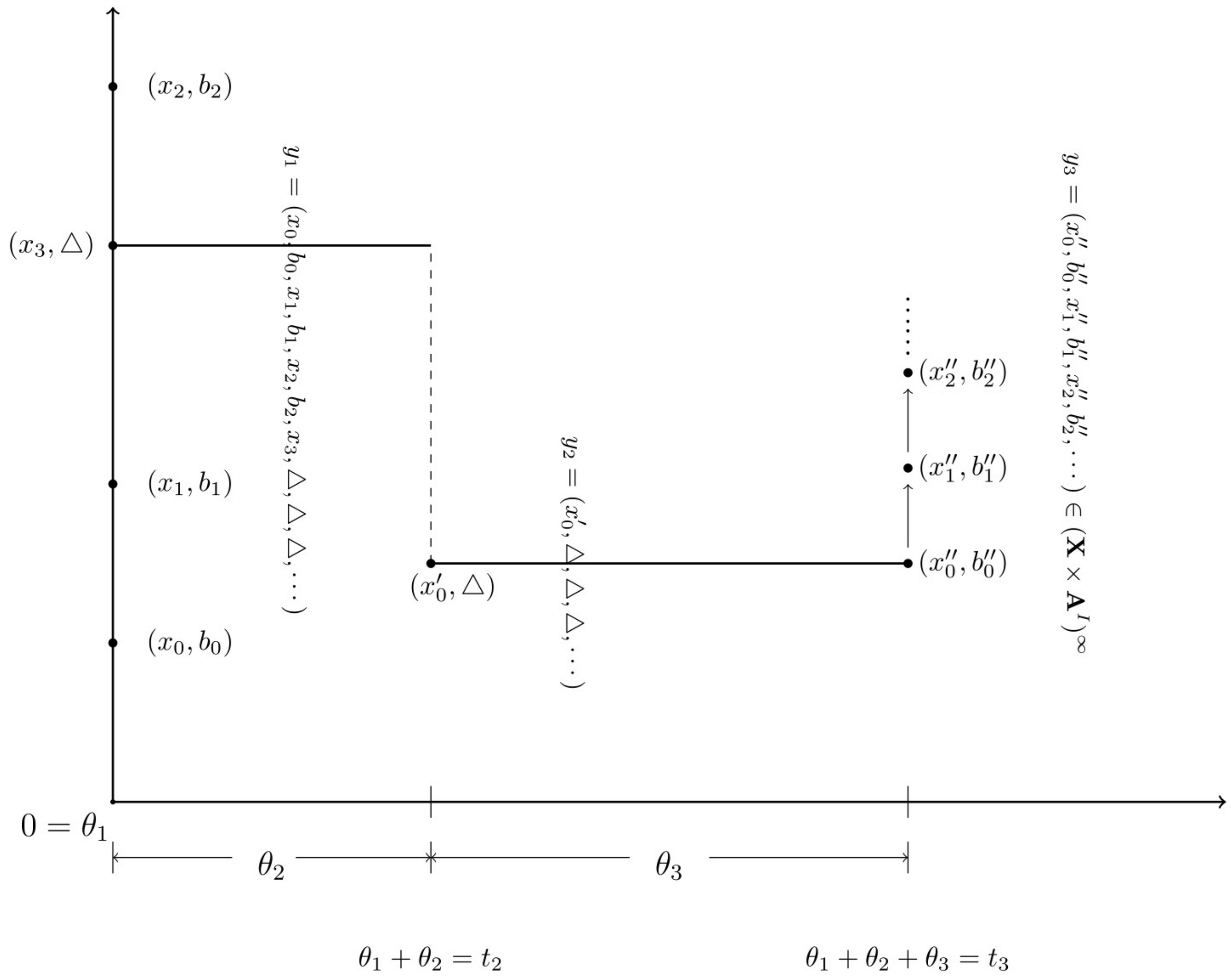
(x_2, b_2)
$y_1 = (x_0, b_0, x_1, b_1, x_2, b_2, x_3, \triangle, \triangle, \triangle, \cdots)$
$(x_3, \triangle)$
$y_3 = (x''_0, b''_0, x''_1, b''_1, x''_2, b''_2, \cdots) \in (\mathbf{X} \times \mathbf{A}^I)^\infty$
(x''_2, b''_2)
$y_2 = (x'_0, \triangle, \triangle, \triangle, \cdots)$
(x''_1, b''_1)
(x_1, b_1)
(x''_0, b''_0)
$(x'_0, \triangle)$
(x_0, b_0)
$0 = \theta_1$
θ_2
θ_3
$\theta_1 + \theta_2 = t_2$
$\theta_1 + \theta_2 + \theta_3 = t_3$

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 2.3, $y_1 \in \mathbf{Y}_3$ 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, $\bar{x}(y_3) = \Delta$.

and

$$G_n^u(dt\times dy|h_n) := \delta_{y_n}(\mathbf{Y})\left\{\Gamma_n^1(dy|h_n)e^{-\Lambda_n^u(\mathbf{X}|h_n,t)}\Phi_n(dt|h_n)\right.$$
$$\left.+\int_{\mathbf{X}}\Phi_n([t,\infty]|h_n)\Gamma_n^0(dy|h_n,t,x)\lambda_n^u(dx|h_n,t)e^{-\Lambda_n^u(\mathbf{X}|h_n,t)}dt\right\} \quad (4)$$

on $(0,\infty)\times\mathbf{Y}$. For each fixed initial state $x_0\in\mathbf{X}$, by the Ionescu-Tulcea theorem, see e.g., Proposition C.10 in [13], there exists a probability $\mathbb{P}_{x_0}^u$ on $(\Omega,\mathcal{F})$ such that the restriction of $\mathbb{P}_{x_0}^u$ to $(\Omega,\mathcal{F}_0)$ is given by

$$\mathbb{P}_{x_0}^u\left((\{y_0\}\times\{0\}\times\Gamma\times((0,\infty]\times\mathbf{Y}_\Delta)^\infty)\cap\Omega\right) \;=\; u_0(\Gamma|x_0) \quad (5)$$

for each $\Gamma\in\mathcal{B}(\mathbf{Y})$; and for each $n\geq 1$, under $\mathbb{P}_{x_0}^u$, 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 $\mathbb{P}_{x_0}^u$ is given by $G_n^u([t,+\infty]\times\mathbf{Y}_\Delta|H_n)$.

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 $x\in\mathbf{X}_\Delta$.) With this notation, we now introduce the performance measure considered in this paper:

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

for each $x\in\mathbf{X}$ and policy $u\in\mathcal{U}$. Here we recall that $T_1=\Theta_1\equiv 0$, see (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

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

More generally, one can compute $\mathbb{E}_x^u\left[e^{C^I(Y_{n+1})}\right]=\mathbb{E}_x^u\left[\mathbb{E}_x^u\left[e^{C^I(Y_{n+1})}|H_n\right]\right]$, where $\mathbb{E}_x^u\left[e^{C^I(Y_{n+1})}|H_n\right]$ can be written out as a similar integral to the case of $n=0$ using the conditional laws (3) and (4).

Let the value function $\mathcal{V}^*$ be denoted by $\mathcal{V}^*(x) := \inf_{u\in\mathcal{U}}\mathcal{V}(x,u)$ for each $x\in\mathbf{X}$. A policy u^* satisfying $\mathcal{V}(x,u^*)=\mathcal{V}^*(x)$ for all $x\in\mathbf{X}$ is called optimal for the gradual-impulse control problem:

$$\text{Minimize over } u\in\mathcal{U}: \mathcal{V}(x,u). \quad (6)$$

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

Definition 2.4 *A policy u is called deterministic stationary if there exist some measurable mappings (φ,ψ,f) on $\boldsymbol{X}$, where $\varphi(x)\in\{0,\infty\}$ for each $x\in\boldsymbol{X}$, ψ and f are $\boldsymbol{A}^I$-valued and $\boldsymbol{A}^G$-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 $u_0(\cdot|x)=\Gamma_n^0(\cdot|h_n,t,x)=\beta^\pi(\cdot|x)$ 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 taking place at some $t\in(0,\infty)$ is always triggered by a natural jump.

3 Optimality result

In this section, we present the main optimality results in this paper. In a nutshell, under natural conditions on the system primitives of the gradual-impulse control problem (6), we show that it can be solved via problem (21) 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 (6) admits a deterministic stationary optimal policy.

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

Condition 3.1 *There exists an $[1,\infty)$-valued continuous function w on $\boldsymbol{X}$ such that $c^G(x,a)+q_x(a)+1 \leq w(x)$ for each $(x,a) \in \boldsymbol{X} \times \boldsymbol{A}^G$.*

If c^G is a continuous function, then the above condition is a consequence of Condition 3.2 below and the Berge theorem, see Proposition 7.32 of [2]. Several statements below do not need the bounding function w in Condition 3.1 to be continuous. In this connection, we also mention that a Borel measurable function w satisfying the inequality in Condition 3.1 always exists, see Lemma 1 of [9] and recall (1).

Recall that $\tilde{\mathbf{A}} = \mathbf{A}^I \cup \mathbf{A}^G$ 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 (6).

Definition 3.1 *The tilde DTMDP model is specified by $\{\boldsymbol{X}, \tilde{\boldsymbol{A}}, \tilde{Q}, \tilde{l}\}$, where $\boldsymbol{X}$ and $\tilde{\boldsymbol{A}}$ are its state and action spaces, and its transition probability $\tilde{Q}$ on $\boldsymbol{X}$ given $\boldsymbol{X} \times \tilde{\boldsymbol{A}}$ 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 $a \in \boldsymbol{A}^G$, and $\tilde{Q}(dy|x,b) := Q(dy|x,b)$, $\tilde{l}(x,b,y) := c^I(x,b,y)$ for all $b \in \boldsymbol{A}^I$.*

For the solvability of problem (21) for the tilde DTMDP model, we impose the following compactness-continuity condition.

Condition 3.2 *The functions c^I and c^G are lower semicontinuous on $\boldsymbol{X} \times \boldsymbol{A}^I \times \boldsymbol{X}$ and $\boldsymbol{X} \times \boldsymbol{A}^G$, resp., and for each bounded continuous function g on $\boldsymbol{X}$, $\int_{\boldsymbol{X}} g(y)Q(dy|x,b)$ and $\int_{\boldsymbol{X}} g(y)\tilde{q}(dy|x,a)$ are continuous in $(x,b) \in \boldsymbol{X} \times \boldsymbol{A}^I$ and $(x,a) \in \boldsymbol{X} \times \boldsymbol{A}^G$, resp.. (Recall also that $\boldsymbol{A}^G$ and $\boldsymbol{A}^I$ are compact.)*

Under Conditions 3.1 and 3.2, one can easily check that the tilde DTMDP model is semicontinuous, so that the value function W^* for problem (21) of the tilde DTMDP model is lower semicontinuous, and there exists an optimal deterministic stationary strategy for it, see Proposition A.1(e). We collect these observations in the next statement for future reference.

Proposition 3.1 *Suppose Conditions 3.1 and 3.2 are satisfied. Then the value function W^* of problem (21) for the tilde DTMDP model is the minimal $[1,\infty]$-valued lower semicontinuous function satisfying*

$$V(x) = \inf_{\tilde{a} \in \tilde{\boldsymbol{A}}} \left\{ \int_{\boldsymbol{X}} e^{\tilde{l}(x,\tilde{a},y)} V(y) \tilde{Q}(dy|x,\tilde{a}) \right\}, \ x \in \boldsymbol{X}. \tag{7}$$

(W^* is also the minimal $[1,\infty]$-valued lower semicontinuous solution to the optimality inequality obtained by replacing in (7) "$=$" with "$\geq$".) A pair of measurable mappings (ψ^*, f^*) from $\boldsymbol{X}$ to $\boldsymbol{A}^I$ and $\boldsymbol{A}^G$, respectively, is a deterministic stationary optimal strategy for problem (21) of the tilde DTMDP model if and only if, for all $x \in \boldsymbol{X}$ there is some x-dependent $\tilde{a}^* \in \tilde{\boldsymbol{A}}$ such that

$$\begin{aligned} & \int_{\boldsymbol{X}} e^{\tilde{l}(x,\tilde{a}^*,y)} W^*(y) \tilde{Q}(dy|x,\tilde{a}^*) = \inf_{\tilde{a} \in \tilde{\boldsymbol{A}}} \left\{ \int_{\boldsymbol{X}} e^{\tilde{l}(x,\tilde{a},y)} W^*(y) \tilde{Q}(dy|x,\tilde{a}) \right\} \\ = & \int_{\boldsymbol{X}} e^{\tilde{l}(x,\psi^*(x),y)} W^*(y) \tilde{Q}(dy|x,\psi^*(x)) I\{\tilde{a}^* \in \boldsymbol{A}^I\} \\ & + \int_{\boldsymbol{X}} e^{\tilde{l}(x,f^*(x),y)} W^*(y) \tilde{Q}(dy|x,f^*(x)) I\{\tilde{a}^* \in \boldsymbol{A}^G\}. \end{aligned} \tag{8}$$

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 $\mathbf{X}$

$$\begin{aligned}\mathbf{X}^G(g) &:= \Big\{x \in \mathbf{X} : \infty > g(x) = \inf_{a\in\mathbf{A}^G}\Big\{\int_{\mathbf{X}} g(y)\tilde{q}(dy|x,a)\\ &\qquad -(q_x(a) - c^G(x,a))g(x)\Big\}\Big\},\end{aligned} \tag{9}$$

and $\mathbf{X}^I(g) := \Big\{x \in \mathbf{X} :\ g(x) = \inf_{b\in\mathbf{A}^I}\Big\{\int_{\mathbf{X}} g(y)e^{c^I(x,b,y)}Q(dy|x,b)\Big\}\Big\}$. Proposition A.1 asserts that, without imposing Condition 3.2, W^* is universally measurable so that the integrals $\int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,a)$ and $\int_{\mathbf{X}} W^*(y)e^{c^I(x,b,y)}Q(dy|x,b)$ are defined.

Theorem 3.1 *Suppose Conditions 3.1 and 3.2 are satisfied. Then the following assertions hold.*
(a) The value function W^* of problem (21) for the tilde DTMDP model coincides with $\mathcal{V}^*$.
(b) $\boldsymbol{X}\setminus \boldsymbol{X}^I(W^*) \subseteq \boldsymbol{X}^G(W^*)$.
(c) There is a deterministic stationary optimal policy for the gradual-impulse control problem (6), which can be obtained as follows. For each pair (ψ^*, f^*) of measurable mappings satisfying (8) (and there exists such a pair by Proposition 3.1), the deterministic stationary policy (φ, ψ^*, f^*) is optimal, where $\varphi(x) = \infty$ for all $x \in \boldsymbol{X}\setminus \boldsymbol{X}^I(W^*)$ and $\varphi(x) = 0$ for all $x \in \boldsymbol{X}^I(W^*)$.

The proofs of this and the other statements in this section are postponed to Section 5.

According to Theorem 3.1, roughly speaking, if the current state is in $\mathbf{X}^G(W^*)$, then it is optimal not to apply impulse until the next natural jump; and if the current state is in $\mathbf{X}^I(W^*)$, then it is optimal to apply immediately an impulse. Also, equation (7) is the optimality equation for the gradual-impulse control problem (6). It can be written out in an equivalent form that does not involve the function w, which might be more convenient sometimes.

Corollary 3.1 *Suppose Conditions 3.1 and 3.2 are satisfied. Then the following assertions hold.*
(a) $\mathcal{V}^*$ is the minimal $[1,\infty]$-valued lower semicontinuous function on $\boldsymbol{X}$ satisfying

$$\begin{aligned}&\inf_{a\in\boldsymbol{A}^G}\Big\{\int_{\boldsymbol{X}} \mathcal{V}^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))\mathcal{V}^*(x)\Big\} \geq 0,\\ &\forall\ x \in \boldsymbol{X}^*(\mathcal{V}^*) := \{x \in \boldsymbol{X} :\ \mathcal{V}^*(x) < \infty\}\end{aligned} \tag{10}$$

and

$$\mathcal{V}^*(x) \leq \inf_{b\in\boldsymbol{A}^I}\Big\{\int_{\boldsymbol{X}} e^{c^I(x,b,y)}\mathcal{V}^*(y)Q(dy|x,b)\Big\},\ x \in \boldsymbol{X}, \tag{11}$$

whereas at each $x \in \boldsymbol{X}$, the inequality in either (10) or (11) holds with equality.
(b) A pair (ψ^*, f^*) of measurable mappings satisfies (8) if and only if

$$\begin{aligned}&\inf_{a\in\boldsymbol{A}^G}\Big\{\int_{\boldsymbol{X}} \mathcal{V}^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))\mathcal{V}^*(x)\Big\}\\ =\ &\int_{\boldsymbol{X}} \mathcal{V}^*(y)\tilde{q}(dy|x,f^*(x)) - (q_x(f^*(x)) - c^G(x,f^*(x)))\mathcal{V}^*(x)\end{aligned}$$

for each $x \in \boldsymbol{X}^G(\mathcal{V}^*)$, and

$$\begin{aligned}\inf_{b\in\boldsymbol{A}^I}\Big\{\int_{\boldsymbol{X}} e^{c^I(x,b,y)}\mathcal{V}^*(y)Q(dy|x,b)\Big\} &= \int_{\boldsymbol{X}} \mathcal{V}^*(y)e^{c^I(x,\psi^*(x),y)}Q(dy|x,\psi^*(x))\\ &\qquad \forall\ x \in \boldsymbol{X}.\end{aligned}$$

(According to Theorem 3.1, (ψ^*, f^*) gives rise to a deterministic stationary optimal policy for the gradual-impulse control problem (6).)

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 3.1 *Let us revisit Example 1.1. 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 $\boldsymbol{A}^I = \{0, 1\}$, with 1 or 0 standing for shooting or not. So the inequalities (10) and (11) for the value function $\mathcal{V}^*$ read:*

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

Suppose $1 - e^C(1-p) > 0$. By Theorem 3.1 and Corollary 3.1, if $\frac{e^C p}{1-e^C(1-p)} > \frac{\mu}{\mu-l} > 0$, then $\mathcal{V}^*(1) = \frac{\mu}{\mu-l}$, and the optimal deterministic stationary policy is to never shoot at the rat; otherwise, $\mathcal{V}^*(1) = \frac{e^C p}{1-e^C(1-p)} = E[e^{CZ}]$ 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 introduced in the next section.

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 this 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 paper 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 coordinate records the system state 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 coordinate records the time in the gradual-impulse control problem elapsed between two consecutive states as recorded in the second coordinate.

The hat DTMDP is with a more complicated action space as compared with the original gradual-impulse control problem. To describe the action space of the hat DTMDP model, let us recall some known and general facts as follows. Let $\mathcal{R}$ be the collection of $\mathcal{P}(\mathbf{A}^G)$-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. Recall that $\mathcal{P}(\mathbf{A}^G)$ stands for the space of probability measures on $(\mathbf{A}^G, \mathcal{B}(\mathbf{A}^G))$. We endow $\mathcal{P}(\mathbf{A}^G)$ with its weak topology (generated by bounded continuous functions on $\mathbf{A}^G$) and the Borel σ-algebra, so that $\mathcal{P}(\mathbf{A}^G)$ is a Borel space, see Chapter 7 of [2]. It is known, see Lemma 1 of [23], 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 $(0,\infty)\times\mathcal{P}(\mathbf{A}^G)$, is a Borel space. Recall that $\mathbf{A}^I$ and $\mathbf{A}^G$ are compact Borel spaces. Then, according to Section 43 of [6], the space $\mathcal{R}$ is a compact metrizable space, endowed with the Young topology, which is the coarsest topology with respect to which, the mapping $\rho=(\rho_t(da))\in\mathcal{R}\to\int_0^\infty\int_{\mathbf{A}^G}g(t,a)\rho_t(da)dt$ is continuous for each function g on $(0,\infty)\times\mathbf{A}^G$ satisfying that (a) for each $t\in(0,\infty)$, $g(t,\cdot)$ is continuous on $\mathbf{A}^G$; (b) for each $a\in\mathbf{A}^G$, $g(\cdot,a)$ is measurable on $(0,\infty)$; and (c) $\int_0^\infty\sup_{a\in\mathbf{A}^G}|g(t,a)|dt<\infty$. Below we shall use, without special reference, the following notation. If μ is a measure on a Borel space $(X,\mathcal{B}(X))$, then $f(\mu):=\int_X f(x)\mu(dx)$ for each measurable function f on $(X,\mathcal{B}(X))$, provided that the integral is well defined.

4.1 Primitives of the hat DTMDP model $\{\hat{\mathbf{X}},\hat{\mathbf{A}},p,l\}$

The state space of the hat DTMDP model is $\hat{\mathbf{X}}:=\{(\infty,x_\infty)\}\cup[0,\infty)\times\mathbf{X}$, 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, $\mathbf{X}$, $\mathbf{A}^I$ 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 $\hat{\mathbf{X}}$ and action $\hat{a}=(c,b,\rho)\in\hat{\mathbf{A}}$,

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

for each state $(\theta,x)\in[0,\infty)\times\mathbf{X}$, and $\int_{\hat{\mathbf{X}}}g(t,y)p(dt\times dy|(\infty,x_\infty),\hat{a}):=g(\infty,x_\infty)$.

It is known, see e.g., [5, 10], that for each bounded measurable function g on $\hat{\mathbf{X}}$, the above expressions are indeed measurable on $\hat{\mathbf{X}}\times\hat{\mathbf{A}}$, and the same also concerns the cost function l on $\hat{\mathbf{X}}\times\hat{\mathbf{A}}\times\hat{\mathbf{X}}$ defined as follows:

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

for each $(\theta,x),\hat{a},(t,y))\in\hat{\mathbf{X}}\times\hat{\mathbf{A}}\times\hat{\mathbf{X}}$, 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.

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

$$\text{Minimize over }\sigma\text{: }\hat{\mathbb{E}}^\sigma_{(\theta,x)}\left[e^{\sum_{n=0}^\infty l(\hat{X}_n,\hat{A}_n,\hat{X}_{n+1})}\right]=:V((\theta,x),\sigma)\qquad(12)$$

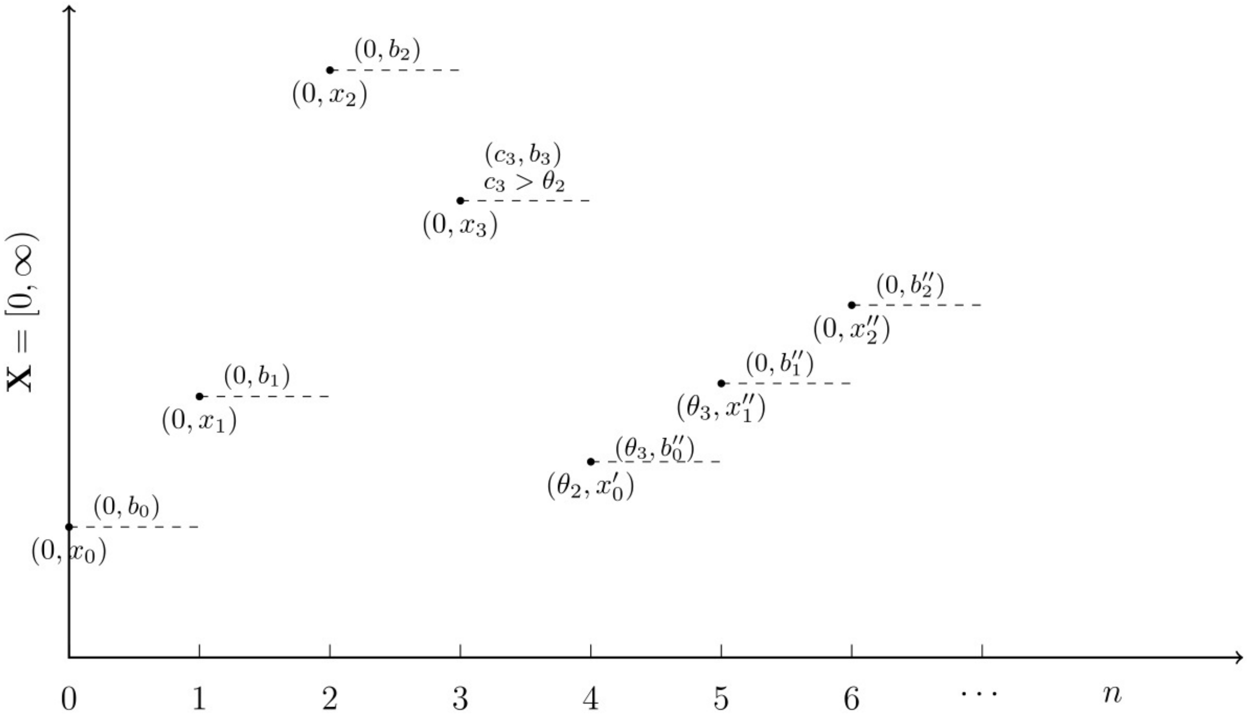

Figure 2: The realization of the state process in the hat DTMDP model corresponding to sample path in the gradual-impulse control problem in Figure 1. The time index is discrete from $\{0, 1, \dots\}$. The realizations of the components $\{(C_n, B_n)\}_{n=0}^{\infty}$ in the action process $\{\hat{A}_n\}_{n=0}^{\infty}$ 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.

where $\{\hat{X}_n\}_{n=0}^{\infty}$ 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. We denote by V^* the value function of this optimal control problem, i.e., $V^*(\theta,x) := \inf_{\sigma} \hat{\mathbb{E}}_{\hat{x}}^{\sigma}\left[e^{\sum_{n=0}^{\infty} l(\hat{X}_n,\hat{A}_n,\hat{X}_{n+1})}\right]$ for each $\hat{x} = (\theta,x) \in \hat{\mathbf{X}}$, where the infimum is over all strategies. Clearly, $V^*(\infty,x_\infty) = 1$. It will be seen in Lemma 5.1 that V^* depends on (θ,x) only through x, and a strategy σ is optimal if $V((0,x),\sigma) = V^*(x)$ for each $x \in \mathbf{X}$. Below, when the context is clear, we often consider the restriction of V^* on $\mathbf{X}$ but still use the same notation.

Let $\hat{h}_n = ((\theta_0,x_0),(c_0,b_0,\rho_0),(\theta_1,x_1),(c_1,b_1,\rho_1),(\theta_2,x_2),\dots,(\theta_n,x_n))$ be the generic notation for an n-history in the hat DTMDP model. A strategy in the hat DTMDP model is a sequence $\sigma = \{\sigma_n\}_{n=0}^{\infty}$, where for each $n \ge 0$, $\sigma_n(d\hat{a}|\hat{h}_n)$ 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 gradual-impulse control problem described in the previous section. To relate the gradual-impulse control problem (6) and the hat DTMDP problem (12), see Proposition 4.1 below, we introduce the following class of strategies in the hat DTMDP model.

Definition 4.1 *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 $[0,\infty] \times \boldsymbol{A}^I$ 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 $\sigma = \{\sigma_n\}_{n=0}^{\infty}$ 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 $\varphi_n(dc|\hat{h}_n)$. 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 $\varphi_n(dc|\hat{h}_n)$ as $\varphi_n(dc|x_n)$.

4.2 Relation between gradual-impulse control and hat DTMDP problems

Each policy u as introduced in Definition 2.3 induces a strategy $\{(\varphi_n,\psi_n,F^n)\}_{n=0}^{\infty}$ in the hat DTMDP model as follows, where we only need consider $x_n \in \mathbf{X}$, as the definition of the strategies at $x_n = x_\infty$ is immaterial, and can be arbitrary. For each $m \ge 1$, and $h_m \in \mathbf{H}_m$, there exists a strategy $\pi^{\Gamma^1_m,h_m} = \{\pi_n^{\Gamma^1_m,h_m}\}_{n=0}^{\infty}$ in the intervention DTMDP model such that $\Gamma^1_m(dy|h_m) = \beta^{\pi^{\Gamma^1_m,h_m}}(dy|\bar{x}(y_m))$. Similarly, for each $x \in \mathbf{X}, t > 0$, there exists a strategy $\pi^{\Gamma^0_m,h_m,t,x} = \{\pi_n^{\Gamma^0_m,h_m,t,x}\}_{n=0}^{\infty}$ in the intervention DTMDP model such that $\Gamma^0_m(dy|h_m,t,x) = \beta^{\pi^{\Gamma^0_m,h_m,t,x}}(dy|x)$. Finally, there is a strategy $\pi^{u_0} = \{(\pi_n^{u_0})\}_{n=0}^{\infty}$ in the intervention DTMDP model satisfying $u_0(dy|x) = \beta^{\pi^{u_0}}(dy|x)$ for each $x \in \mathbf{X}$.

Consider the case of $n = 0$. Then we define $\varphi_0(\{0\}|\theta,x) := 1 - \pi_0^{u_0}(\{\Delta\}|x)$; $\varphi_0(dc|\theta,x) :=$

$\pi_0^{u_0}(\{\Delta\}|x)\Phi_1(dc|(x,\Delta,\Delta,\dots),0,(x,\Delta,\Delta,\dots))$ on $(0,\infty]$;

$$
\begin{aligned}
\psi_0(db|\theta,x,c) &:= \frac{\pi_0^{u_0}(db|x)}{1-\pi_0^{u_0}(\{\Delta\}|x)}I\{c=0\} \\
&+I\{c>0\}\frac{\pi_0^{\Gamma_1^1,((x,\Delta,\dots),0,(x,\Delta,\dots))}(db|x)}{1-\pi_0^{\Gamma_1^1,((x,\Delta,\dots),0,(x,\Delta,\dots))}(\{\Delta\}|x)} \\
&= \frac{\pi_0^{u_0}(db|x)}{1-\pi_0^{u_0}(\{\Delta\}|x)}I\{c=0\}+I\{c>0\}\pi_0^{\Gamma_1^1,((x,\Delta,\dots),0,(x,\Delta,\dots))}(db|x);
\end{aligned}
$$

and $F^0(\theta,x)_t(da):=\Pi_1(da|(x,\Delta,\Delta,\dots),0,(x,\Delta,\Delta,\dots),t)$, 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 2.3. Also concerning the definition of $\psi_0(db|\theta,x,c)$, note that if the denominator $1-\pi_0^{u_0}(\{\Delta\}|x)=0$, 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 $x\in\mathbf{X}$, 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. 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 $\Gamma_1\subseteq(0,\infty]$,

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

which is equal to

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

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

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

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 $\hat{h}_n$ be the n-history in the hat DTMDP model. If $\{1 \leq i \leq n : \ \theta_i > 0\} = \emptyset$, then we define $\varphi_n(\{0\}|\hat{h}_n) := 1 - \pi_n^{u_0}(\{\Delta\}|x_0, b_0, \dots, b_{n-1}, x_n)$;

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

and $F^n(\hat{h}_n)_t(da) := \Pi_1(da|y_0, 0, (x_0, b_0, \dots, x_n, \Delta, \Delta, \dots), t)$. Recall the notation that was introduced earlier: $y_0 = (x_0, \Delta, \Delta, \dots)$.

If $\{1 \leq i \leq n : \ \theta_i > 0\} \neq \emptyset$, then let $m(\hat{h}_n) := \#\{1 \leq i \leq n : \ \theta_i > 0\}$, and $l(\hat{h}_n) := \max\{1 \leq i \leq n : \theta_i > 0\}$. 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 $i \geq 1$. 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 $y_0 = (x_0, \Delta, \Delta, \dots)$; $y_1 = (x_0, b_0, x_1, b_1, \dots, x_{\tau_1 - 1}, \Delta, \Delta, \dots)$; if $\theta_{\tau_1} = c_{\tau_1 - 1}$, then $y_2 = (x_{\tau_1 - 1}, b_{\tau_1 - 1}, x_{\tau_1}, b_{\tau_1}, \dots, x_{\tau_2 - 1}, \Delta, \Delta, \dots)$, if $\theta_{\tau_1} < c_{\tau_1 - 1}$, then

$$y_2 = (x_{\tau_1}, b_{\tau_1}, \dots, x_{\tau_2 - 1}, \Delta, \Delta, \dots); \ \dots;$$

if $\theta_{\tau_{m-1}} = c_{\tau_{m-1}-1}$, then $y_m = (x_{\tau_{m-1}-1}, \dots, x_{\tau_m - 1}, \Delta, \Delta, \dots)$, and if $\theta_{\tau_{m-1}} < c_{\tau_{m-1}-1}$, then

$$y_m = (x_{\tau_{m-1}}, \dots, x_{\tau_m - 1}, \Delta, \Delta, \dots).$$

For example, if

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

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

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

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

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

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

Proof. One can verify

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

for each $n \geq 1$. 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 4.1 *A deterministic stationary policy say u^D identified by (φ, ψ, f) as in Definition 2.4 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 $t \geq 0$, vice versa. It is evident that $\mathcal{V}(x, u^D) = V(x, \sigma^D)$ for each $x \in \boldsymbol{X}$. Thus, if the hat DTMDP problem (12) has an optimal strategy in this form of $\sigma^D = (\varphi, \psi, F)$, then the previous discussions lead to $\mathcal{V}^*(x) = V^*(x)$, and that the deterministic stationary policy u^D associated with σ^D is optimal for the gradual-impulse control problem (6).*

To end this section, note that Condition 3.2 does not imply that the hat DTMDP model is semicontinuous, which is defined in the appendix. In fact, the transition probability p, in general, does not satisfy the weak continuity condition, even under Condition 3.2. The simplest example is as follows.

Example 4.1 *Suppose $q_x(a) \equiv 0$, and $\boldsymbol{A}^G$ and $\boldsymbol{A}^I$ 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 $\hat{\boldsymbol{X}}$: $g(t, x) \equiv 1$ for each $(t, x) \in [0, \infty) \times \boldsymbol{X}$, and $g(\infty, x_\infty) = 0$. Then $\int_{\hat{\boldsymbol{X}}} g(t, y) p(dt \times dy|(\theta, x), \hat{a}_n) = \int_{\boldsymbol{X}} g(c_n, y) Q(dy|x, b) = 1$ for each $n \geq 1$, whereas $\int_{\hat{\boldsymbol{X}}} g(t, y) p(dt \times dy|(\theta, x), (\infty, b, \rho)) = g(\infty, x_\infty) = 0 \neq 1$.*

One can also construct examples, where the transition probability p is not continuous with respect to $\rho \in \mathcal{R}$.

5 Proof of the main statements

In this section, we prove the results stated in Section 3. This is based on the investigation of problem (12) for the hat DTMDP model described in Section 4. In this section, unless specified otherwise, V^* is understood as the value function of problem (12) for the hat DTMDP model. The main properties concerning V^* are summarized in the next statement.

Proposition 5.1 *(a) V^* is a $[1, \infty]$-valued lower semianalytic function on $\boldsymbol{X}$ satisfying*

$$\inf_{a \in \boldsymbol{A}^G} \left\{ \int_{\boldsymbol{X}} V^*(y) \tilde{q}(dy|x, a) - (q_x(a) - c^G(x, a)) V^*(x) \right\} \geq 0, \qquad (13)$$
$$\forall\, x \in \boldsymbol{X}^*(V^*) := \{x \in \boldsymbol{X} : \ V^*(x) < \infty\}$$

and

$$V^*(x) \leq \inf_{b \in \boldsymbol{A}^I} \left\{ \int_{\boldsymbol{X}} e^{c^I(x,b,y)} V^*(y) Q(dy|x, b) \right\}, \ x \in \boldsymbol{X}, \qquad (14)$$

whereas at each $x \in \boldsymbol{X}$, the inequality in either (13) or (14) holds with equality.
(b) $\boldsymbol{X} \setminus \boldsymbol{X}^I \subseteq \boldsymbol{X}^G$, where $\boldsymbol{X}^G := \boldsymbol{X}^G(V^*)$, see (9), and $\boldsymbol{X}^I := \boldsymbol{X}^I(V^*)$. (Lemma 5.1 below asserts that V^* is universally measurable so that the integrals $\int_{\boldsymbol{X}} V^*(y) \tilde{q}(dy|x, a)$ and $\int_{\boldsymbol{X}} V^*(y) e^{c^I(x,b,y)} Q(dy|x, b)$ are defined.)

Proof. See Lemmas 5.1, 5.3 and 5.4 below. □

Lemma 5.1 *(a) The value function V^* depends on the state (θ, x) only through the second coordinate, and thus we write $V^*(x)$ instead of $V^*(\theta, x)$. The function V^* is an $[1, \infty]$-valued lower semianalytic* *function satisfying*

$$\begin{aligned} V(x) &= \inf_{\hat{a}\in\hat{\boldsymbol{A}}}\left\{\int_0^c \int_{\boldsymbol{X}} V(y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds}dt \right. \\ &\quad +I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds} \\ &\quad \left. +I\{c<\infty\}e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\boldsymbol{X}} V(y)e^{c^I(x,b,y)}Q(dy|x,b)\right\},\ x\in\boldsymbol{X}; \\ V(x_\infty) &= 1, \end{aligned} \tag{15}$$

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

$$\begin{aligned} V(x) &\geq \inf_{\hat{a}\in\hat{\boldsymbol{A}}}\left\{\int_0^c \int_{\boldsymbol{X}} V(y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds}dt \right. \\ &\quad +I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds} \\ &\quad \left. +I\{c<\infty\}e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\boldsymbol{X}} V(y)e^{c^I(x,b,y)}Q(dy|x,b)\right\},\ x\in\boldsymbol{X}; \\ V(x_\infty) &= 1. \end{aligned} \tag{16}$$

(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 (12).*
(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 (15) with* V^* *replacing* V*, for each* $x\in\boldsymbol{X}$*.*
(d) For each $x\in\boldsymbol{X}$*,* $V^*(x)=\inf_{\pi\in\Pi^U}V(x,\pi)$*, where* Π^U *is 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

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

where the inequality can be replaced by equality, follows from Proposition A.1. The existence of an ϵ-optimal deterministic Markov universally measurable strategy follows from Proposition A.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. [8, 25]. Now assertions (a,b) follow. Finally, the last two assertions follow from Proposition A.1. □

Lemma 5.2 *The function in* $t\in[0,\infty)$ *defined by*

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

is increasing, for each $x\in\boldsymbol{X}$ *and* $\rho\in\mathcal{R}$*.*

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

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

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,

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

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

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

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

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

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 5.1, 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

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

where the first inequality holds because of the last assertion of Lemma 5.1. Since the expression in the last bracket is nonnegative and finite, and $\epsilon > 0$ was arbitrarily fixed, we see that $V^*(x) \le \int_0^{t_2-t_1} e^{-\int_0^\tau (q_x(\tilde{\rho}_s)-c^G(x,\tilde{\rho}_s))ds} \int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,\tilde{\rho}_\tau)d\tau + e^{-\int_0^{t_2-t_1}(q_x(\tilde{\rho}_s)-c^G(x,\tilde{\rho}_s))ds} V^*(x)$, as desired. □

Lemma 5.3 *Relations (13) and (14) hold. (Recall from Lemma 5.1 that V^* is universally measurable.)*

Proof. Let $x \in \mathbf{X}$ be fixed. Inequality (14) immediately follows from Lemma 5.1, if on the right hand side of (15) 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 (13) as follows. Suppose $V^*(x) < \infty$. Let $a \in \mathbf{A}^G$ be arbitrarily fixed. If $\int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,a) = \infty$, then trivially, $\int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))V^*(x) \ge 0$. Consider the case when $\int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,a) < \infty$. Let $t > 0$ be arbitrarily fixed. Then $\int_0^t e^{-\tau(q_x(a)-c^G(x,a))} \int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,a)d\tau + e^{-t(q_x(a)-c^G(x,a))}V^*(x)$ is finite. Upon differentiating it with respect to t and applying the fundamental theorem of calculus, we see

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

where the inequality follows from Lemma 5.2. Thus, $\int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))V^*(x) \ge 0$. Since $a \in \mathbf{A}^G$ was arbitrarily fixed, we see that (13) holds. □

Lemma 5.4 *For each $x \in \mathbf{X}$, the inequality in either (13) or (14) holds with equality.*

Proof. Let $x \in \mathbf{X}$ be fixed. If the equality in (14) holds at this point, then there is nothing to prove. Suppose the strict inequality holds in (14). Then necessarily $V^*(x) < \infty$. The objective is to show that, in this case, (13) holds with equality. For the infimum in (15) with V^* replacing V, it suffices to consider $c > 0$, because (14) holds with strict inequality at the fixed point $x \in \mathbf{X}$ here. Let $\epsilon > 0$ be fixed, and $(c^*, b^*, \rho^*) \in \hat{\mathbf{A}}$ be such that

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

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

Consider case (a). Then

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

where the second inequality holds because of (14), and the last inequality holds because of Lemma 5.2. Thus, as $\epsilon>0$ was arbitrarily fixed,

$$\begin{aligned}V^*(x) &= \inf_{\rho\in\mathcal{R}}\left\{\int_0^{c^*} e^{-\int_0^t(q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,\rho_t)dt\right.\\ &\left.+e^{-\int_0^{c^*}(q_x(\rho_s)-c^G(x,\rho_s))ds}V^*(x)\right\}. \end{aligned}\tag{17}$$

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

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

and

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

where the last inequality holds because of (13). Since $\int_0^{c^*} e^{-\overline{q}_x s}ds>0$ and $\delta>0$ was arbitrarily fixed, we see that (13) holds with equality.

Now consider case (b). Then

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

One can show that for each $t \in [0, \infty)$,

$$\begin{aligned} V^*(x) \;&=\; \inf_{\rho\in\mathcal{R}} \left\{ \int_0^t e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds} \int_{\mathbf{X}} V^*(y)\tilde{q}(dy|x,\rho_t)dt \right. \\ &\quad \left. + e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds} V^*(x) \right\}. \end{aligned} \tag{18}$$

The details are 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

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

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

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

where the second inequality is by Lemma 5.1(a), which in particular, asserts that V^* satisfies (15), and the last inequality is by Lemma 5.2. Since $\epsilon > 0$ and $\delta > 0$ were arbitrarily fixed, the above implies (18). Comparing (18) with (17), we see that case (b) is reduced to case (a). □

Lemma 5.5 *Let w be a measurable $[1, \infty)$-valued function satisfying the inequality in Condition 3.1, whose existence is guaranteed as mentioned in the paragraph below Condition 3.1. Consider the transition probability $\tilde{p}(dy|x, a)$ on $\mathcal{B}(\boldsymbol{X})$ given $(x, a) \in \boldsymbol{X} \times \boldsymbol{A}^G$ defined by $\tilde{p}(\Gamma|x, a) := \frac{q(\Gamma|x,a)}{w(x)} + \delta_x(dy)$, $\forall\, \Gamma \in \mathcal{B}(\boldsymbol{X})$, $(x, a) \in \boldsymbol{X} \times \boldsymbol{A}^G$. Then an $[1, \infty]$-valued lower semianalytic function V^* (here the notation V^* does not necessarily mean the value function) satisfies (13) and (14), and for each $x \in \boldsymbol{X}$, either (13) or (14) holds with equality, if and only if it satisfies (14), for each $x \in \boldsymbol{X}$*

$$V^*(x) \leq \inf_{a\in\boldsymbol{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x, a)} \int_{\boldsymbol{X}} V^*(y)\tilde{p}(dy|x, a) \right\}, \tag{19}$$

and either (14) or (19) holds with equality, i.e.,

$$\begin{aligned} V^*(x) \;&=\; \min \left\{ \inf_{a\in\boldsymbol{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x, a)} \int_{\boldsymbol{X}} V^*(y)\tilde{p}(dy|x, a) \right\}, \right. \\ &\quad \left. \inf_{b\in\boldsymbol{A}^I} \left\{ \int_{\boldsymbol{X}} V^*(y) e^{c^I(x,b,y)} Q(dy|x, b) \right\} \right\}. \end{aligned} \tag{20}$$

(Note that (19) automatically holds with equality at $x \in \mathbf{X} \setminus \mathbf{X}^*(V^*) := \{x \in \mathbf{X} : V^*(x) = \infty\}$. Also note that the function w in the previous lemma does not need be continuous.)

Proof of Lemma 5.5. "Only if" part. Consider an $[1, \infty]$-valued lower semianalytic function V^* that satisfies (13) and (14), and for each $x \in \mathbf{X}$, either (13) or (14) holds with equality. For $x \in \mathbf{X}^*(V^*) = \{x \in \mathbf{X} : V^*(x) < \infty\}$, (13) implies for each $a \in \mathbf{A}^G$ that $0 \le c^G(x,a)V^*(x) + \int_{\mathbf{X}} V^*(y)q(dy|x,a) = (c^G(x,a) - w(x))V^*(x) + w(x)\int_{\mathbf{X}} V^*(y)\tilde{p}(dy|x,a)$, and so

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

i.e., (19) holds. Let $x \in \mathbf{X}^*(V^*)$ be a point, where (13) holds with equality. Let us verify at this point $x \in \mathbf{X}^*(V^*)$, (19) also holds with equality. For each $\epsilon > 0$, there is some $a_\epsilon \in \mathbf{A}^G$ such that $\epsilon \ge c^G(x,a_\epsilon)V^*(x) + \int_{\mathbf{X}} V^*(y)q(dy|x,a_\epsilon)$ so that

$$\begin{aligned} & V^*(x) + \epsilon \ge V^*(x) + \frac{\epsilon}{w(x) - c^G(x,a_\epsilon)} \ge V^*(x) \\ & + \frac{c^G(x,a_\epsilon)V^*(x) + \int_{\mathbf{X}} V^*(y)q(dy|x,a_\epsilon)}{w(x) - c^G(x,a_\epsilon)} = \frac{w(x)}{w(x) - c^G(x,a_\epsilon)} \int_{\mathbf{X}} \tilde{p}(dy|x,a_\epsilon)V^*(y) \\ \ge\ & \inf_{a \in \mathbf{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x,a)} \int_{\mathbf{X}} V^*(y)\tilde{p}(dy|x,a) \right\}, \end{aligned}$$

and thus $V^*(x) \ge \inf_{a \in \mathbf{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x,a)} \int_{\mathbf{X}} V^*(y)\tilde{p}(dy|x,a) \right\}$. The opposite direction of this inequality was seen earlier, and so (19) holds with equality at this point. This completes the "Only if" part. The argument for the "If" part is the same, and omitted. □

Remark 5.1 *Consider the function V^* in the previous statement. By inspecting the above proof we see the following fact: a pair of measurable mappings ψ^* and f^* from $\boldsymbol{X}$ to $\boldsymbol{A}^I$ and $\boldsymbol{A}^G$ satisfy*

$$\begin{aligned} & \frac{w(x)}{w(x) - c^G(x,f^*(x))} \int_{\boldsymbol{X}} V^*(y)\tilde{p}(dy|x,f^*(x)) \\ =\ & \inf_{a \in \boldsymbol{A}^G} \left\{ \frac{w(x)}{w(x) - c^G(x,a)} \int_{\boldsymbol{X}} V^*(y)\tilde{p}(dy|x,a) \right\} \end{aligned}$$

for each $x \in \boldsymbol{X}$, at which (19) holds with equality, and

$$\int_{\boldsymbol{X}} e^{c^I(x,\psi^*(x),y)} V^*(y)Q(dy|x,\psi^*(x)) = \inf_{b \in \boldsymbol{A}^I} \left\{ \int_{\boldsymbol{X}} e^{c^I(x,b,y)} V^*(y)Q(dy|x,b) \right\} \ \forall\ x \in \boldsymbol{X},$$

if and only if

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

for each $x \in \boldsymbol{X}$, at which the left hand side equals zero, and

$$\int_{\boldsymbol{X}} e^{c^I(x,\psi^*(x),y)} V^*(y)Q(dy|x,\psi^*(x)) = \inf_{b \in \boldsymbol{A}^I} \left\{ \int_{\boldsymbol{X}} e^{c^I(x,b,y)} V^*(y)Q(dy|x,b) \right\} \ \forall\ x \in \boldsymbol{X}.$$

Lemma 5.6 *Suppose Conditions 3.1 and 3.2 are satisfied. Then $W^*(x) = V^*(x)$ for each $x \in \boldsymbol{X}$.*

Proof. According to Proposition A.1(a,b), the value function W^* for the tilde model is the minimal $[1,\infty]$-valued lower semianalytic function satisfying (7) as well as the inequality obtained by replacing the equality in (7) by "$\geq$". Let us verify that $W^* = V^*$ as follows. According to Lemmas 5.3, 5.4 and 5.5, the value function V^* is a $[1,\infty]$-valued lower semianalytic function satisfying (7), c.f. (20). Therefore, $W^* \leq V^*$ pointwise.

For the opposite direction of this inequality, let $x \in \mathbf{X}$ be fixed. It suffices to show that W^* satisfies (16) at the point x. Then, since the point $x \in \mathbf{X}$ was arbitrarily fixed, one could apply Lemma 5.1 to obtain $V^* \leq W^*$ pointwise.

Recall that, as observed in the beginning of this proof, W^* satisfies (20). By Lemma 5.5, it satisfies (13) and (14), one of which holds with equality at this point x. If (14) holds with equality for W^* at x, then

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

and thus (16) is satisfied by W^* at x, as required. Now suppose (13) holds with equality for W^* at x. It suffices to consider $W^*(x) < \infty$, for otherwise, (16) automatically holds for W^* at x. According to Remark 5.1 after Lemma 5.5 and because the tilde model is semicontinuous, there is some $a^* \in \mathbf{A}^G$ satisfying

$$\begin{aligned} &\int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,a^*) - (q_x(a^*) - c^G(x,a^*))W^*(x)\\ =\ &\inf_{a\in\mathbf{A}^G}\left\{\int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,a) - (q_x(a) - c^G(x,a))W^*(x)\right\} = 0, \end{aligned}$$

and hence $\int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,a^*) = (q_x(a^*) - c^G(x,a^*))W^*(x)$. 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,

$$\begin{aligned} &W^*(x) \geq 1 = \int_0^\infty\int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,a^*)e^{-\int_0^t(q_x(a^*)-c^G(x,a^*))ds}dt\\ &+e^{-\int_0^\infty q_x(a^*)ds}e^{\int_0^\infty c^G(x,a^*)ds} \geq \inf_{\hat{a}\in\hat{\mathbf{A}}}\left\{\int_0^c\int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,\rho_t)\right.\\ &e^{-\int_0^t(q_x(\rho_s)-c^G(x,\rho_s))ds}dt + I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds}\\ &\left.+I\{c<\infty\}e^{-\int_0^c(q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\mathbf{X}} W^*(y)e^{c^I(x,b,y)}Q(dy|x,b)\right\}. \end{aligned}$$

That is, (16) is satisfied by W^* at x, as desired. Finally, if $c^G(x,a^*) < q_x(a^*)$, then

$$\begin{aligned}
&\inf_{\hat{a}\in\hat{\mathbf{A}}}\left\{\int_0^c \int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,\rho_t)e^{-\int_0^t (q_x(\rho_s)-c^G(x,\rho_s))ds}dt\right.\\
&+I\{c=\infty\}e^{-\int_0^\infty q_x(\rho_s)ds}e^{\int_0^\infty c^G(x,\rho_s)ds}\\
&\left.+I\{c<\infty\}e^{-\int_0^c (q_x(\rho_s)-c^G(x,\rho_s))ds}\int_{\mathbf{X}} W^*(y)e^{c^I(x,b,y)}Q(dy|x,b)\right\}\\
\le\ & \int_0^\infty \int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,a^*)e^{-\int_0^t (q_x(a^*)-c^G(x,a^*))ds}dt + e^{-\int_0^\infty q_x(a^*)ds}e^{\int_0^\infty c^G(x,a^*)ds}\\
=\ & \frac{\int_{\mathbf{X}} W^*(y)\tilde{q}(dy|x,a^*)}{q_x(a^*)-c^G(x,a^*)} + 0 = W^*(x),
\end{aligned}$$

as requested. Thus, W^* satisfies (16). Consequently, $W^* = V^*$ on $\mathbf{X}$, as required. □

Proof of Theorem 3.1. Part (b) was seen in the proof of Lemma 5.4.

Consider the pair of measurable mappings (ψ^*, f^*) from Proposition 3.1. Recall that $W^* = V^*$ on $\mathbf{X}$ by Lemma 5.6. Keeping in mind Remark 5.1, an inspection of the proof of Lemma 5.6 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

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

for each $x \in \mathbf{X}$. By Theorem 5.1, this deterministic stationary strategy $(\varphi(x), \psi^*(x), t \to \delta_{f^*(x)}(da)) \in \hat{\mathbf{A}}$ is optimal for problem (12) for the hat DTMDP model. This and Remark 4.1 imply that $V^* = \mathcal{V}^*$ on $\mathbf{X}$ and part (c). By Lemma 5.6, we see now $\mathcal{V}^* = W^*$ on $\mathbf{X}$, and thus part (a) holds. □

Proof of Corollary 3.1. This corollary follows at once from Theorem 3.1, Lemma 5.5 and Remark 5.1. □

6 Conclusion

In this paper we investigated the gradual-impulse control problem of CTMDPs with a rigorous and general construction, which allows the consideration of quite a large class of control policies, and is not restricted to the form of performance measures under consideration. Possible future research thus includes the investigation of gradual-impulse control problem of CTMDPs with other performance measures (such as the long run average cost in the risk-neutral as well as risk-sensitive setups.)

A Relevant results about DTMDPs

In this appendix, we present the relevant facts about DTMDPs. The proofs of the presented statements can be found in [15] or [25]. Standard description of a DTMDP can be found in e.g., [13, 20]. The notations used in this section are independent of the previous sections.

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

Condition A.1 *(a) The function $l(x,a,y)$ is lower semicontinuous in $(x,a,y) \in \boldsymbol{X} \times \boldsymbol{A} \times \boldsymbol{X}$.*
(b) For each bounded continuous function f on $\boldsymbol{X}$, $\int_{\boldsymbol{X}} f(y)p(dy|x,a)$ is continuous in $(x,a) \in \boldsymbol{X} \times \boldsymbol{A}$.
(c) The space $\boldsymbol{A}$ is a compact Borel space.

The DTMDP model $\{\mathbf{X}, \mathbf{A}, p,\mathrm{l}\}$ is called semicontinuous if it satisfies Condition A.1.

Let us denote for each $n = 1, 2, \ldots, \infty$, $\mathbf{H}_n := \mathbf{X} \times (\mathbf{A} \times \mathbf{X})^n$ and $\mathbf{H}_0 := \mathbf{X}$. A strategy $\sigma = (\sigma_n)_{n=0}^{\infty}$ in the DTMDP is given by a sequence of stochastic kernels $\sigma_n(da|h_n)$ on $\mathcal{B}(\mathbf{A})$ from $h_n \in \mathbf{H}_n$ for $n = 0, 1, 2, \ldots$. 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 $\mathbf{X}$. 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 $\mathbf{X}$ to $\mathbf{A}$. Let Σ be the space of strategies, and Σ_{DM} be the space of all deterministic Markov strategies for the DTMDP.

Let the controlled and controlling process be denoted by $\{Y_n\}_{n=0}^{\infty}$ and $\{A_n\}_{n=0}^{\infty}$. 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 $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$, by the Ionescu-Tulcea theorem, c.f., [13, 20], one can construct a probability measure $\mathbb{P}_\nu^\sigma$ on $(\mathbf{H}_\infty, \mathcal{B}(\mathbf{H}_\infty))$ such that

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

As usual, equalities involving conditional expectations and probabilities are understood in the almost sure sense.

The probability measure $\mathbb{P}_\nu^\sigma$ is called a strategic measure (of the strategy σ) in the DTMDP model $\{\mathbf{X}, \mathbf{A}, p,\mathrm{l}\}$ (with the initial distribution ν). The expectation taken with respect to $\mathbb{P}_\nu^\sigma$ is denoted by $\mathbb{E}_\nu^\sigma$. When ν is concentrated on the singleton $\{x\}$, $\mathbb{P}_\nu^\sigma$ and $\mathbb{E}_\nu^\sigma$ are written as $\mathbb{P}_x^\sigma$ and $\mathbb{E}_x^\sigma$.

Consider the optimal control problem

$$\text{Minimize over } \sigma: \quad \mathbb{E}_x^\sigma\left[e^{\sum_{n=0}^{\infty} l(Y_n, A_n, Y_{n+1})}\right] =: \mathbf{V}(x, \sigma),\ x \in \mathbf{X}. \tag{21}$$

We denote the value function of problem (21) by $\mathbf{V}^*$. Then a strategy σ^* is called optimal for problem (21) if $\mathbf{V}(x, \sigma^*) = \mathbf{V}^*(x)$ for each $x \in \mathbf{X}$. For a constant $\epsilon > 0$, a strategy is called ϵ-optimal for problem (21) if $\mathbf{V}(x, \sigma^*) \leq \mathbf{V}^*(x) + \epsilon$ for each $x \in \mathbf{X}$.

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 [2] 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 [25] in the next proposition.

Proposition A.1 *(a) The value function $\boldsymbol{V}^*$ is the minimal $[1, \infty]$-valued lower semianalytic solution to*

$$\boldsymbol{V}(x) = \inf_{a \in \boldsymbol{A}} \left\{ \int_{\boldsymbol{X}} e^{l(x,a,y)}\, \boldsymbol{V}(y) p(dy|x, a) \right\},\ x \in \boldsymbol{X}. \tag{22}$$

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

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

then $\boldsymbol{U}(x) \geq \boldsymbol{V}^*(x)$ for each $x \in \boldsymbol{X}$.
(c) Let φ be a deterministic stationary strategy for the DTMDP model $\{\boldsymbol{X}, \boldsymbol{A}, p, l\}$. If

$$\boldsymbol{V}^*(x) = \int_{\boldsymbol{X}} e^{l(x,\varphi(x),y)}\, \boldsymbol{V}^*(y)p(dy|x,\varphi(x)),\ \forall\ x \in \boldsymbol{X}, \tag{23}$$

then $\boldsymbol{V}^*(x) = \boldsymbol{V}(x,\varphi)$ for each $x \in \boldsymbol{X}$.
(d) $\boldsymbol{V}^*(x) = \inf_{\sigma\in\Sigma^U} \boldsymbol{V}(x,\sigma)$, 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 (21).
(e) Suppose Condition A.1 is satisfied. Then the value function $\boldsymbol{V}^*$ is the minimal $[1,\infty]$-valued lower semicontinuous solution to (22). Moreover, there exists a deterministic stationary strategy φ satisfying (23), and so in particular, there exists a deterministic stationary optimal strategy for problem (21).

Part (d) of the above statement follows from the proof of Proposition 3.2 of [25], whereas all the other parts are according to Propositions 3.1, 3.4 and 3.7 therein.

Acknowledgements

We thank the editors and referees for comments and remarks that improve significantly the readability of this paper. This work is supported by the Royal Society (grant number IE160503), and the Daiwa Anglo-Japanese Foundation (UK) (grant reference 4530/12801).

References

[1] Bäuerle, N. and Popp, A. (2018). Risk-sensitive stopping problems for continuous-time Markov chains. *Stochastics* **90,** 411–431.

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

[3] Costa, O. and Davis, M. (1989). Impulsive control of piecewise-deterministic processes. *Math. Control Signals Systems* **2,** 187–206.

[4] Costa, O. and Raymundo, C. (2000). Impulse and continuous control of piecewise deterministic Markov processes. *Stochastics* **70,** 75–107.

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

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

[7] Dufour, F. and Piunovskiy, A. (2015). Impulsive control for continuous-time Markov decision processes. *Adv. Appl. Probab.* **47,** 106–127.

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

[9] Feinberg, E., Mandava, M. and Shiryaev, A. (2017). Kolmogorov's equations for jump Markov processes with unbounded jump rates. *Ann. Oper. Res.* (accepted). DOI 10.1007/s10479-017-2538-8.

[10] 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.

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

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

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

[14] Hordijk, A. and van der Duyn Shouten, F. (1984). Discretization and weak convergence in Markov decision drift processes. *Math. Oper. Res.* **9,** 121–141.

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

[16] Kumar, S. and Pal, C. (2013). Risk-sensitive control of pure jump process on countable space with near monotone cost. *Appl. Math. Optim.* **68,** 311-331.

[17] Miller, A., Miller, B. and Stepanyan, K. (2018). Simultaneous impulse and continuous control of a Markov chain in continuous time. *Automation and Remote Control* **81,** 469–482.

[18] 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.

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

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

[21] van der Duyn Schouten, F. (1983). *Markov Decision Processes with Continuous Time Parameter.* Mathematisch Centrum, Amsterdam.

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

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

[24] Yushkevich, A. (1988). Bellman inequalities in Markov decision dterministic drift processes. *Stochastics* **23,** 25–77.

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