On reducing a constrained gradual-impulsive control problem for a jump Markov model to a model with gradual control only

Alexey Piunovskiy* and Yi Zhang †

Abstract: In this paper we consider a gradual-impulsive control problem for continuous-time Markov decision processes (CTMDPs) with total cost criteria and constraints. We develop a simple and useful method, which reduces the concerned problem to a standard CTMDP problem with gradual control only. This allows us to derive straightforwardly and under a minimal set of conditions, the optimality results (sufficient classes of control policies, as well as the existence of stationary optimal policies) for the original constrained gradual-impulsive control problem.

Keywords: Continuous-time Markov decision processes. Impulse-gradual control. Reduction method. Problem with constraints.

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

1 Introduction

In this paper, we study the gradual-impulsive control problems for continuous-time Markov decision processes (CTMDPs). The objective is to minimize the expected total cost, subject to the constraints that several other expected total costs cannot be too large.

A large portion of the previous literature on CTMDPs, see e.g., the monographs [18, 32], focuses on the gradual control problems, in which, the decision maker can influence the underlying system dynamics through the control of the local characteristics (transition intensity and post jump distribution). This is in contrast with the impulsive control problem, where the decision maker can directly and instantaneously change the state of the process under control. A typical application is in reliability, where the maintenance (or replacement) activity can change the status of the facility immediately. Likewise, there are wide and natural applications of impulsive control of CTMDPs in queueing systems, epidemiology, etc, see e.g. [8, 26, 27].

A gradual-impulsive control problem allows the controller to control the underlying system both impulsively and gradually. Gradual-impulsive control problem has been studied intensively since 1970s, with the pioneering work [2], where the process under control is a diffusion process. One of the first works on the gradual-impulsive control of CTMDPs (with deterministic drift between two consecutive jumps) seems to be [33], which was later extended to piecewise deterministic processes (PDPs) with boundary jumps in [4, 5, 6, 13, 17]. Very often in the literature, one concentrates on policies that apply at maximum only one impulse at each single time moment, because multiple impulses at a single time moment may lead to nonstandard trajectories with multiple values at a single time moment. A rigorous construction of the gradual-impulsive control problem allowing one to consider multiple simultaneous impulsive controls was given in [35], where the motivations for

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

considering multiple impulses at a single time moment were provided, especially when impulses have random effects. An alternative description was given in the more recent work [11, 13]. A more recent work on gradual-impulsive control of CTMDPs is [25]. These aforementioned works considered the impulsive or gradual-impulsive control problem with a single objective, and their investigations were based on the Bellman optimality equation.

The present paper considers the gradual-impulsive control problems for CTMDPs with multiple objectives (one being minimized while the others are subject to constraints). To the best of our knowledge, this type of constrained gradual-impulsive control problems was only studied in [12], dealing with discounted problems. The convex analytic approach was developed in [12], making essential use of versions of Kolmogorov equations. This method is an extension of the one in [28] for gradual control problems, and requires certain regularity conditions imposed on the system parameters. For example, the transition and cost rates were assumed to be bounded in [12]. This restriction would exclude many natural applications, such as the controlled $M/M/\infty$ system with holding cost. In contrast to that, the present paper does not require the boundedness on the transition and cost rates. To this end, the present paper suggests a different method from [12], which is more transparent on the one hand, and requires only a minimal set of conditions on the system parameters on the other hand. The main contributions of the present paper lie in this, and are elaborated as follows:

(a) We show that the constrained gradual-impulsive control problem for CTMDPs can be reduced to an equivalent standard CTMDP problem (with only gradual controls), and obtain sufficient class of policies in a simple form. This is done under a minimal set of conditions in the sense that the claimed result may not hold without these conditions, as demonstrated by an example.

(b) As a demonstration of the application of the above result, by referring to the known fact about standard CTMDPs (with gradual controls only) obtained in e.g., [19], we show, under a natural set of compactness-continuity conditions, that there exists an optimal stationary policy for the constrained gradual-impulsive control problem for CTMDPs with nonnegative (gradual) cost rates and impulse cost functions. No boundedness condition is needed on the growth of the transition and cost rates and impulse cost functions.

Let us say a few words on the novelty of the reduction method in this paper. The method of reducing a gradual-impulsive control problem to an equivalent gradual control problem also appeared in [7], where the authors studied a single-objective gradual-impulse control of PDPs. The key idea in [7] is that if the state is extended, then the impulsive control of the original process can be viewed as a boundary control in the new model with only gradual control (as well as the boundary control, which occurs as soon as the controlled process hits the boundary of the state space.) To serve this idea, the authors (a) extended the state space such that the state of the new model is much more complicated (a six-tuple) than the original one; and (b) were restricted to a special class of deterministic control policies that do not apply multiple simultaneous impulses. The reduction method in the present paper is different, and does not follow the idea of [7]. Roughly speaking, the idea here is to view the successive states after a sequence of impulses at the same time moment "horizontally" instead of "vertically". To the best of our knowledge, the current idea had not been employed before. Consequently, in the present paper, (a) the reduced CTMDP model (with only gradual control and without boundary control) is quite simple and has the same state space as the original model; and (b) we consider general class of control policies that allow multiple simultaneous impulses. Despite we do not follow it in this paper, let us mention that another popular method for investigating gradual-impulsive control of CTMDPs is by time-discretization, see [21, 31].

Finally, the term of "impulse control problem with constraint" also appeared in [24], see also [23], but with a different meaning. In [23, 24], the constraint was imposed on when an impulse could be

applied. Here, the constraints come from the multiple objectives. The interested reader can also find references on impulse control problem of other classes of processes in [23, 24].

The rest of this paper is organized as follows. In Section 2, we describe the gradual-impulsive control model as well as the standard CTMDP model with gradual control only. In Section 3, we present the main optimality results. Their proofs are postponed to Section 4. Some further remarks are given in Section 5. This paper ends with a conclusion in Section 6.

2 Model descriptions

We fist introduce some notations, definitions and facts to be used below, often without special reference. A Borel space is a Borel measurable subset of a complete separable metric space. Suppose $\mathbf{X}$ is a Borel space endowed with its Borel σ-algebra $\mathcal{B}(\mathbf{X})$. Let $\mathcal{P}(\mathbf{X})$ stand for the space of probability measures on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$. We denote by $\mathcal{R}(\mathbf{X})$ the collection of $\mathcal{P}(\mathbf{X})$-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. Throughout this text, unless stated otherwise, by measurable we mean Borel measurable.

2.1 Gradual-impulsive control model

We describe the primitives of the gradual-impulsive control 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, there are a family of cost rates and functions $\{c_i^G, c_i^I\}_{i=0}^J$, with J being a fixed positive integer, representing the number of constraints in the concerned optimal control problem to be described below, see (3). For each $i \in \{0,1,\ldots,J\}$, c_i^G and c_i^I are $[-\infty,\infty]$-valued measurable functions on $\mathbf{X} \times \mathbf{A}^G$ and $\mathbf{X} \times \mathbf{A}^I$, respectively.

Remark 2.1 *It is without loss of generality to assume $\boldsymbol{A}^G$ and $\boldsymbol{A}^I$ as two disjoint measurable subsets of a Borel space $\boldsymbol{A}$ such that $\boldsymbol{A} = \boldsymbol{A}^G \bigcup \boldsymbol{A}^I$, for otherwise, one can consider $\boldsymbol{A}^G \times \{G\}$ instead of $\boldsymbol{A}^G$ and $\boldsymbol{A}^I \times \{I\}$ instead of $\boldsymbol{A}^I$ and $\boldsymbol{A} = \boldsymbol{A}^G \times \{G\} \bigcup \boldsymbol{A}^I \times \{I\}$.*

The description of the system dynamics in the gradual-impulsive control problem is as follows. Assume $q_x(a) > 0$ for each $x \in \mathbf{X}$ and $a \in \mathbf{A}^G$ for simplicity. At the initial time 0 with the initial state x_0, the decision maker selects the triple $(\hat{c}_0, \hat{b}_0, \rho^0)$ with $\hat{c}_0 \in [0,\infty]$, $\hat{b}_0 \in \mathbf{A}^I$, and $\rho^0 = \{\rho_t^0(da)\}_{t \in (0,\infty)} \in \mathcal{R}(\mathbf{A}^G)$. Then, the time until the next natural jump follows the nonstationary exponential distribution with the rate function $\int_{\mathbf{A}^G} q_{x_0}(a)\rho_t^0(da) =: q_{x_0}(\rho_t^0)$. Here and below, if $\rho \in \mathcal{R}(\mathbf{A}^G)$, then $q_x(\rho_t) := \int_{\mathbf{A}^G} q_x(a)\rho_t(da)$ and $\tilde{q}(dy|x,\rho_t) := \int_{\mathbf{A}^G} \tilde{q}(dy|x,a)\rho_t(da)$. If by time $\hat{c}_0$, there

is no occurrence of a natural jump, then the first sojourn time is $\hat{c}_0$, at which, the impulsive action $\hat{b}_0 \in \mathbf{A}^I$ is applied, and the next state X_1 follows the distribution $Q(dy|x_0, \hat{b}_0)$. If the first natural jump happens before $\hat{c}_0$, say at t_1, then the first sojourn time is t_1, and the next state X_1 follows the distribution $\frac{\tilde{q}(dy|x_0,\rho^0_{t_1})}{q_{x_0}(\rho^0_{t_1})}$. Except for the initial one, a decision epoch occurs immediately after a sojourn time. At the next decision epoch, the decision maker selects $(\hat{c}_1, \hat{b}_1, \rho^1)$, and so on. This leads to a natural description of the gradual-impulsive control problem as a discrete-time Markov decision process (DTMDP), which is presented next. This way of describing the gradual-impulsive control problem for a CTMDP is due to Yushkevich [35].

The state space of the DTMDP model corresponding to the gradual-impulsive control problem is $\hat{\mathbf{X}} := \{(\infty, x_\infty)\} \bigcup [0, \infty) \times \mathbf{X}$, where (∞, x_∞) is an isolated point in $\hat{\mathbf{X}}$. The first coordinate represents the previous sojourn time in the gradual-impulsive control problem, and the state of the controlled process in the gradual-impulsive control problem is given in the second coordinate. The inclusion of the first coordinate in the state allows us to consider control policies that select actions depending on the past sojourn times.

The action space of the DTMDP is $\hat{\mathbf{A}} := [0, \infty] \times \mathbf{A}^I \times \mathcal{R}(\mathbf{A}^G)$. Recall that $\mathcal{R}(\mathbf{A}^G)$ is 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, where $\mathcal{P}(\mathbf{A}^G)$ 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 [3]. According to Lemma 3 of [34], each element in $\mathcal{R}(\mathbf{A}^G)$ can be regarded as a stochastic kernel from $(0, \infty)$ to $\mathcal{B}(\mathbf{A}^G)$. According to Lemma 1 of [34], the space $\mathcal{R}(\mathbf{A}^G)$, endowed with the smallest σ-algebra with respect to which the mapping $\rho = (\rho_t(da)) \in \mathcal{R}(\mathbf{A}^G) \to \int_0^\infty e^{-t} g(t, \rho_t) dt$ is measurable for each bounded measurable function g on $(0, \infty) \times \mathcal{P}(\mathbf{A}^G)$, is a Borel space.

The transition probability p in the DTMDP is defined as follows. For each bounded measurable function g on $\hat{\mathbf{X}}$ and action $\hat{a} = (\hat{c}, \hat{b}, \rho) \in \hat{\mathbf{A}}$,

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

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

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

The object p defined above is indeed a stochastic kernel from $\hat{\mathbf{X}} \times \hat{\mathbf{A}}$ to $\mathcal{B}(\hat{\mathbf{X}})$, see Lemma 2 of [34] and its proof therein. Similarly, the cost functions $\{l_i\}_{i=0}^J$ defined below are measurable on $\hat{\mathbf{X}} \times \hat{\mathbf{A}} \times \hat{\mathbf{X}}$:

$$
\begin{aligned}
l_i((\theta, x), \hat{a}, (t, y)) \; &:= \; I\{(\theta, x) \in [0, \infty) \times \mathbf{X}\} \left\{ \int_0^t c_i^G(x, \rho_s) ds + I\{t = \hat{c} < \infty\} c_i^I(x, \hat{b}) \right\} \\
&= \; I\{x \in \mathbf{X}\} \left\{ \int_0^t c_i^G(x, \rho_s) ds + I\{t = \hat{c} < \infty\} c_i^I(x, \hat{b}) \right\}, \qquad (2)
\end{aligned}
$$

for each $i = 0, 1, \ldots, J$ and $((\theta, x), \hat{a}, (t, y)) \in \hat{\mathbf{X}} \times \hat{\mathbf{A}} \times \hat{\mathbf{X}}$. Here the generic notation $\hat{a} = (\hat{c}, \hat{b}, \rho) \in \hat{\mathbf{A}}$ of an action in this DTMDP model has been in use. The interpretation is that the pair $(\hat{c}, \hat{b})$ 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. Without loss of generality, the initial state is $(0, x_0)$, with some $x_0 \in \mathbf{X}$.

Let $\{\hat{X}_n\}_{n=0}^{\infty} = \{(\hat{\Theta}_n, X_n)\}_{n=0}^{\infty}$ and $\{\hat{A}_n\}_{n=0}^{\infty}$ be the controlled and controlling process in this DTMDP model, and $\{(\hat{C}_n, \hat{B}_n)\}_{n=0}^{\infty}$ the coordinate process corresponding to $\{(\hat{c}_n, \hat{b}_n)\}_{n=0}^{\infty}$ in $\{\hat{a}_n\}_{n=0}^{\infty}$.

Next, we define the concerned class of policies in the gradual-impulsive control model.

Definition 2.1 *Consider a sequence of stochastic kernels $\sigma = \{\sigma_n\}_{n=0}^{\infty}$, where for each $n \geq 0$, σ_n is a stochastic kernel on $\mathcal{B}([0, \infty] \times \boldsymbol{A}^I \times \mathcal{R}(\boldsymbol{A}^G))$ given $\hat{h}_n := (\hat{x}_0, (\hat{c}_0, \hat{b}_0), \hat{x}_1, (\hat{c}_1, \hat{b}_1), \ldots, \hat{x}_n)$. According to Proposition 7.27 of [3],*

$$\sigma_n(d\hat{c} \times d\hat{b} \times d\rho|\hat{h}_n) = \sigma_n^{(0)}(d\hat{c} \times d\hat{b}|\hat{h}_n)\sigma_n^{(1)}(d\rho|\hat{h}_n, \hat{c}, \hat{b}),$$

where $\sigma_n^{(0)}$ and $\sigma_n^{(1)}$ are some corresponding stochastic kernels. If for each $n \geq 0$, there is a measurable mapping $\hat{F}_n$ mapping $(\hat{h}_n, \hat{c}, \hat{b})$ to $\mathcal{R}(\boldsymbol{A}^G)$ such that

$$\sigma_n^{(1)}(d\rho|\hat{h}_n, \hat{c}, \hat{b}) = \delta_{\hat{F}_n(\hat{h}_n, \hat{c}, \hat{b})}(d\rho),$$

then we call the sequence $\sigma = \{\sigma_n\}_{n=0}^{\infty}$, which is also identified with $\sigma = \{\sigma_n^{(0)}, \hat{F}_n\}_{n=0}^{\infty}$, a policy for the gradual-impulsive control model. The collection of all policies for the gradual-impulsive CTMDP model is denoted by Σ.

Under a policy σ, having in hand $\hat{h}_n$, the decision maker selects $(\hat{c}_n, \hat{b}_n)$ (possibly randomly), and after that, chooses $\rho^n = \hat{F}_n(\hat{h}_n, \hat{c}_n, \hat{b}_n)$.

Given $\hat{x}_0 = (0, x_0) \in \hat{\mathbf{X}}$ and a policy σ, let $\hat{\mathrm{P}}_{x_0}^{\sigma}$ be the strategic measure in the DTMDP, and $\hat{\mathrm{E}}_{x_0}^{\sigma}$ the corresponding expectation. Then the concerned gradual-impulsive control problem with constraints reads

$$\begin{aligned} &\text{Minimize over } \sigma \in \Sigma : \hat{\mathrm{E}}_{x_0}^{\sigma}\left[\sum_{n=0}^{\infty} l_0(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1})\right] =: \hat{W}_0(x_0, \sigma) \\ &\text{such that} \quad \hat{W}_j(x_0, \sigma) := \hat{\mathrm{E}}_{x_0}^{\sigma}\left[\sum_{n=0}^{\infty} l_j(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1})\right] \leq d_j, \ j = 1, \ldots, J, \end{aligned} \tag{3}$$

where $\{d_j\}_{j=1}^{J} \subset \mathbb{R}^J$ is a fixed vector of constants, x_0 is a fixed element of $\mathbf{X}$, and

$$\hat{\mathrm{E}}_{x_0}^{\sigma}\left[\sum_{n=0}^{\infty} l_i(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1})\right] := \hat{\mathrm{E}}_{x_0}^{\sigma}\left[\sum_{n=0}^{\infty} l_i^+(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1})\right] - \hat{\mathrm{E}}_{x_0}^{\sigma}\left[\sum_{n=0}^{\infty} l_i^-(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1})\right]$$

with $\infty - \infty := \infty$ being adopted here.

We shall obtain the optimality results for this problem in Section 3, by using a novel and simple technique, which reduces the constrained gradual-impulsive control problem to a standard constrained CTMDP problem with gradual control only, which we describe in the next subsection.

2.2 Standard CTMDP model

In a standard CTMDP model, there is only gradual control, which is selected according to purely relaxed policies. Its system primitives are the following objects

$$\mathcal{M}^{GO} := \{\mathbf{X}, \mathbf{A}, q^{GO}, \{c_i^{GO}\}_{i=0}^{J}\}.$$

Here the state and action spaces $\mathbf{X}$ and $\mathbf{A}$ are Borel spaces, q^{GO} is the transition rate from $\mathbf{X}\times\mathbf{A}$ to $\mathcal{B}(\mathbf{X})$, and $\{c_i^{GO}\}_{i=0}^J$ is the collection of measurable functions on $\mathbf{X}\times\mathbf{A}$, representing the cost rates, $J\geq 0$ is a fixed integer. The superscript "GO" abbreviates "gradual only", as the model only allows gradual controls.

In the standard CTMDP model $\mathcal{M}^{GO}$, a decision epoch occurs after each natural jump of the controlled process (except for the initial decision epoch at time zero). At each decision epoch, one selects the relaxed control $\rho\in\mathcal{R}(\mathbf{A})$ until the next decision epoch occurs. We sketch the more rigorous construction as follows. The sample space Ω is taken as the union of $(\mathbf{X}\times(0,\infty))^\infty$ and the collection of sequences in the form $(x_0,\theta_1,x_1,\dots,\theta_{m-1},x_{m-1},\theta_m,x_\infty,\infty,x_\infty,\dots)$, where $m\geq 1$, and $x_\infty\notin\mathbf{X}$ is an isolated point. We endow Ω with the σ-algebra $\mathcal{F}$ obtained as the trace of $\mathcal{B}((\mathbf{X}_\infty\times(0,\infty])^\infty)$ on Ω, where $\mathbf{X}_\infty=\mathbf{X}\bigcup\{x_\infty\}$. The generic notation for an element of Ω is ω. For each $\omega\in\Omega$, define $\theta_0:=0$, $t_n:=\sum_{i=0}^n\theta_i$, $h_n:=(x_0,\theta_1,x_1,\dots,\theta_n,x_n)$ for each $n\geq 0$. The collection of all possible h_n is denoted as $\mathbf{H}_n$ for each $n\geq 0$. Let us put $t_\infty:=\lim_{n\to\infty}t_n$, which exists. When regarded as coordinate variables, we use capital letters Θ_n, T_n, X_n, and H_n corresponding to θ_n,t_n,x_n and h_n. The state process $\{X(t)\}_{t\geq 0}$ is defined by $X(t):=X_n$ if $T_n\leq t<T_{n+1}$ for some $n\geq 0$, and $X(t):=x_\infty$ if $t\geq T_\infty$. As usual, we omit ω whenever the context excludes confusion.

Definition 2.2 *A strategy*[1] $\overline{S}$ *in the standard CTMDP model* $\mathcal{M}^{GO}$ *is the following object:* $\overline{S}=\{\overline{F}_n\}_{n=0}^\infty$, *for each* $n\geq 0$, $\overline{F}_n$ *is a measurable mapping on* $\boldsymbol{H}_n$ *taking values in* $\mathcal{R}(\boldsymbol{A})$.

Remark 2.2 *We put* $q_{x_\infty}^{GO}(a)\equiv 0\equiv q^{GO}(\Gamma|x_\infty,a)$ *for all* $\Gamma\in\mathcal{B}(\boldsymbol{X})$ *and* $c_i^{GO}(x_\infty,a)\equiv 0$ *in what follows.*

Given a strategy $\overline{S}=\{\overline{F}_n\}_{n=0}^\infty$ and initial state $x_0\in\mathbf{X}$, there is a unique probability measure $\mathrm{P}_{x_0}^{\overline{S}}$ on $(\Omega,\mathcal{F})$ such that $\mathrm{P}_{x_0}^{\overline{S}}(X_0\in dx)=\delta_{x_0}(dx)$, and for each $n\geq 1$ and $\Gamma_1\in\mathcal{B}([0,\infty))$, $\Gamma_2\in\mathcal{B}(\mathbf{X})$,

$$\begin{aligned}&\mathrm{P}_{x_0}^{\overline{S}}(\Theta_n\in\Gamma_1,\ X_n\in\Gamma_2|H_{n-1})\\ =\ &\int_{\Gamma_1}e^{-\int_0^s q_{X_{n-1}}^{GO}(\overline{F}_{n-1}(H_{n-1})_t)dt}\tilde{q}^{GO}(\Gamma_2|X_{n-1},\overline{F}_{n-1}(H_{n-1})_s)ds;\\ &\mathrm{P}_{x_0}^{\overline{S}}(\Theta_n=\infty,\ X_n=x_\infty|H_{n-1})=e^{-\int_0^\infty q_{X_{n-1}}^{GO}(\overline{F}_{n-1}(H_{n-1})_t)dt};\end{aligned}$$

and

$$\mathrm{P}_{x_0}^{\overline{S}}(\Theta_n=\infty,\ X_n\in\Gamma_2|H_{n-1})=\mathrm{P}_{x_0}^{\overline{S}}(\Theta_n\in\Gamma_1,\ X_n=x_\infty|H_{n-1})=0.$$

Let the expectation corresponding to $\mathrm{P}_{x_0}^{\overline{S}}$ be denoted as $\mathrm{E}_{x_0}^{\overline{S}}$. We consider the following optimal control problem corresponding to problem (3):

$$\begin{aligned}&\text{Minimize over }\overline{S}:W_0(x_0,\overline{S}):=\mathrm{E}_{x_0}^{\overline{S}}\left[\sum_{n=0}^\infty I\{T_n<\infty\}\int_{T_n}^{T_{n+1}}c_0^{GO}(X_n,\overline{F}_n(H_n)_{t-T_n})dt\right]\\ \text{such that}\quad &W_j(x_0,\overline{S}):=\mathrm{E}_{x_0}^{\overline{S}}\left[\sum_{n=0}^\infty I\{T_n<\infty\}\int_{T_n}^{T_{n+1}}c_j^{GO}(X_n,\overline{F}_n(H_n)_{t-T_n})dt\right]\leq d_j,\\ &j=1,\dots,J,\end{aligned}\tag{4}$$

[1]The term strategy is a synonym of the term policy, but we use it exclusively for models with gradual control only.

where

$$\begin{aligned} & \mathrm{E}_{x_0}^{\overline{S}}\left[I\{T_n<\infty\}\sum_{n=0}^{\infty}\int_{T_n}^{T_{n+1}} c_i^{GO}(X_n,\overline{F}_n(H_n)_{t-T_n})dt\right] \\ := \; & \mathrm{E}_{x_0}^{\overline{S}}\left[\sum_{n=0}^{\infty} I\{T_n<\infty\}\int_{T_n}^{T_{n+1}} c^{GO}{}_i^{+}(X_n,\overline{F}_n(H_n)_{t-T_n})dt\right] \\ & -\mathrm{E}_{x_0}^{\overline{S}}\left[\sum_{n=0}^{\infty} I\{T_n<\infty\}\int_{T_n}^{T_{n+1}} c^{GO}{}_i^{-}(X_n,\overline{F}_n(H_n)_{t-T_n})dt\right], \end{aligned}$$

with $\infty-\infty:=\infty$ being accepted here. Here, the constants J and $\{d_j\}_{j=1}^{J}$ are the same as in problem (3), and we have used the following notation: for each probability measure μ on $\mathcal{B}(\mathbf{X})$ and measurable function f on $\mathbf{X}$, we put $f(\mu):=\int_{\mathbf{X}} f(x)\mu(dx)$ whenever the right hand side is well defined. This notation is only for brevity, and will be used when there is no potential confusion regarding the underlying space $\mathbf{X}$.

We may also write

$$W_i(x_0,\overline{S}) \;=\; \mathrm{E}_{x_0}^{\overline{S}}\left[\sum_{n=0}^{\infty}\int_{0}^{\Theta_{n+1}} c^{GO}{}_i^{+}(X_n,\overline{F}_n(H_n)_t)dt\right]-\mathrm{E}_{x_0}^{\overline{S}}\left[\sum_{n=0}^{\infty}\int_{0}^{\Theta_{n+1}} c^{GO}{}_i^{-}(X_n,\overline{F}_n(H_n)_t)dt\right].$$

In the literature of CTMDPs with gradual controls, it is the standard model $\mathcal{M}^{GO}$ that has been primarily investigated, see e.g., the monographs [18, 22, 32].

The main result in this paper is that the gradual-impulsive control problem can be reduced to the gradual control problem for a standard CTMDP model, and the latter problem was better studied. (In particular, the standard CTMDP problem (4) was studied in [19, 29].) We will justify this by comparing the performance measures $\{\hat{W}_i(x_0,\sigma)\}_{i=0}^{J}$ and $\{W_i(x_0,\overline{S})\}_{i=0}^{J}$. In doing so, we also obtain a sufficient class of policies for solving the impulsive-gradual control problem (3).

3 Main statements

In this section we present the main results concerning the gradual-impulsive control model described in Subsection 2.1.

Condition 3.1 *$Q(\{x\}|x,a)=0$ for each $(x,a)\in \boldsymbol{X}\times\boldsymbol{A}^I$*

This condition is not restrictive because one can always extend the state space $\mathbf{X}$ to $\mathbf{X}\times\{0,1\}$, say, where the second component does not affect any primitives, but switches from 0 to 1 and back from 1 to 0 at every transition moment associated with the impulsive control.

Throughout this section, we consider the following standard CTMDP model $\mathcal{M}^{GO}$, where

$$\begin{aligned} & \mathbf{A}:=\mathbf{A}^I\bigcup\mathbf{A}^G;\; q^{GO}(dy|x,a):=q(dy|x,a),\;\forall\;(x,a)\in\mathbf{X}\times\mathbf{A}^G; \\ & \tilde{q}^{GO}(dy|x,a):=Q(dy|x,a),\; q_x^{GO}(a):=1,\;\forall\;(x,a)\in\mathbf{X}\times\mathbf{A}^I; \\ & c_i^{GO}(x,a):=c_i^G(x,a),\;\forall\;(x,a)\in\mathbf{X}\times\mathbf{A}^G;\; c_i^{GO}(x,a):=c_i^I(x,a),\;\forall\;(x,a)\in\mathbf{X}\times\mathbf{A}^I. \end{aligned} \tag{5}$$

Condition 3.1 guarantees that q^{GO} defined in the above is indeed a transition rate.

Theorem 3.1 *Suppose Condition 3.1 is satisfied, and there is some $\epsilon > 0$ such that $q_x(a) \geq \epsilon > 0$ for all $x \in \boldsymbol{X}$ and $a \in \boldsymbol{A}^G$. For each strategy $\overline{S} = \{\overline{F}_n\}_{n=0}^{\infty}$ in the standard CTMDP model $\mathcal{M}^{GO}$, there is some policy $\sigma = \{\sigma_n^{(0)}, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ in the gradual-impulsive CTMDP model such that*

$$\hat{W}_i(x_0, \sigma) = W_i(x_0, \overline{S})$$

for each $i = 0, 1, \ldots, J$. Moreover, one can take the required policy $\sigma = \{\sigma_n^{(0)}, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ in such a form that, for all $n \geq 0$,

$$\begin{aligned}
&\sigma_n^{(0)}(\{\infty\} \times d\hat{b}|\hat{h}_n) = \sigma_n^{(0)}(\{\infty\} \times d\hat{b}|x_n) = \mu_n(x_n)\bar{\varphi}_n(d\hat{b}|x_n), \\
&\sigma_n^{(0)}(\{0\} \times d\hat{b}|x_n) = (1 - \mu_n(x_n))\bar{\varphi}_n(d\hat{b}|x_n), \\
&\hat{F}_n(\hat{h}_n)_t(da) \equiv \hat{F}_n(x_n)(da),
\end{aligned} \tag{6}$$

where μ_n and $\hat{F}_n$ are $[0,1]$- and $\mathcal{R}(\boldsymbol{A}^G)$-valued measurable mappings on $\boldsymbol{X}$, and $\overline{\varphi}_n$ is a stochastic kernel. The first and last equality indicate that the dependence on $\hat{h}_n$ is only through x_n, the right hand side of the last equality does not depend on t, which has thus been omitted from the subscript. (Note that $\sigma_n^{(0)}(\{0, \infty\} \times \boldsymbol{A}^I|x_n) \equiv 1$.)

The proofs of all the theorems in this section are postponed to Section 4 below.

The opposite direction of the previous theorem also holds.

Theorem 3.2 *Suppose Condition 3.1 is satisfied. For each policy $\sigma \in \Sigma$ in the gradual-impulsive CTMDP model, there is some strategy $\bar{S} = \{\overline{F}_n\}_{n=0}^{\infty}$ in the standard CTMDP model such that $\hat{W}_i(x_0, \sigma) = W_i(x_0, \bar{S})$ for each $i = 0, 1, \ldots, J$.*

Remark 3.1 *The previous two theorems reduce the gradual-impulsive control problem (3) to a standard CTMDP problem (4) with gradual control only. This gives rise to a method of studying the gradual-impulsive control problem (3), which we demonstrate in the proof of Theorem 3.3 below, where it is also pointed out how to produce an optimal policy for the gradual-impulsive control problem (3) from an optimal strategy for the standard CTMDP problem (4) .*

Another straightforward consequence of Theorems 3.1 and 3.2 is the following one concerning the sufficient class of policies for solving the gradual-impulsive control problem (3). Its proof is obvious and thus omitted.

Corollary 3.1 *Suppose Condition 3.1 is satisfied, and there is some $\epsilon > 0$ such that $q_x(a) \geq \epsilon > 0$ for all $x \in \boldsymbol{X}$ and $a \in \boldsymbol{A}^G$. Then for each given policy $\sigma' \in \Sigma$, there exists a policy $\sigma = \{\sigma_n^{(0)}, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ in the form of (6) in the gradual-impulsive control problem (3) such that $\hat{W}_i(x_0, \sigma') = \hat{W}_i(x_0, \sigma)$ for each $i = 0, 1, \ldots, J$. In particular, if there is an optimal policy for the gradual-impulsive control problem (3), then there exists an optimal one in the form of (6).*

The sufficiency result obtained in the above corollary does not require any compactness-continuity conditions. It will be further strengthened below if we impose such conditions.

A particular example of gradual-impulsive control problem is the optimal stopping problem, where the process is pushed to a cemetery once it is stopped. In the first glance, this was excluded if in our model, there is some $\epsilon > 0$ such that $q_x(a) \geq \epsilon > 0$ for all $x \in \mathbf{X}$ and $a \in \mathbf{A}^G$. However, the cemetery can be replicated with a loop between two states, which jumps from one to the other without any cost, as in the next example. This example also demonstrates that the assertion of Corollary 3.1 may not hold in general if $\inf_{a \in \mathbf{A}^G} q_x(a) = 0$ for some $x \in \mathbf{X}$.

Example 3.1 *Let* $\boldsymbol{X} = \{0,1,2,3\}$, $\boldsymbol{A}^G = (0,\infty)$, *and* $\boldsymbol{A}^I = \{0\}$. *Consider the case of* $J = 1$ *with* $c_1^G(x,a) = c_1^I(x,0) \equiv 0 = d_1$ *so that all the policies are feasible for problem (3). Let us only focus on* $\hat{W}_0$ *with* c_0^G *and* c_0^I *satisfying*

$$\begin{aligned}
&c_0^G(0,a) \equiv -1,\ c_0^I(0,0) = \infty,\\
&0 = c_0^G(1,a) = c_0^G(2,a) = c_0^G(3,a) = c_0^I(1,0) = c_0^I(2,0) = c_0^I(3,0).
\end{aligned}$$

Let

$$\begin{aligned}
&q_0(a) = e^{-a} = q(\{1\}|0,a),\ \forall\ a \in \boldsymbol{A}^G;\ q_x(a) = 1,\ \forall\ x \in \{1,2,3\},\\
&1 \equiv q(\{2\}|1,a) = q(\{3\}|2,a) = q(\{2\}|3,a),\\
&1 = Q(\{1\}|0,0) = Q(\{2\}|1,0) = Q(\{3\}|2,0) = Q(\{2\}|3,0).
\end{aligned}$$

Finally, let $x_0 = 0$. *Consider any policy* $\sigma^* = \{\sigma^{*(0)}_n, \hat{F}^*_n\}_{n=0}^{\infty}$ *satisfying* $\hat{F}^*_0(0)_t(da) = \delta_t(da)$, *and* $\sigma^{*(0)}_0(\{\infty\} \times \{0\}|0) = 1$. *Then*

$$\hat{W}_0(0,\sigma^*) = -\infty,$$

because under this policy, no impulse is applied before the next natural jump, whereas the time duration $\hat{\Theta}_1$ *until the first natural jump is infinite with probability* e^{-1}, *as* $\hat{\mathrm{P}}^{\sigma^*}_0(\hat{\Theta}_1 > s) = e^{-\int_0^s e^{-t}dt} = e^{e^{-s}-1}$. *In particular,* σ^* *is an optimal policy for the gradual-impulsive control problem (3).*

On the other hand, for each policy $\sigma = \{\sigma^{(0)}_n, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ *in the form of (6),*

$$\hat{W}_0(0,\sigma) \geq -\frac{1}{\int_{\boldsymbol{A}^G} e^{-a}\hat{F}_0(0)(da)} > -\infty,$$

where the denominator in the last fraction is strictly positive. This means, all the policies $\sigma = \{\sigma^{(0)}_n, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ *in the form of (6) are not optimal, and in particular, the statement of Corollary 3.1 does not hold. The conditions imposed in Corollary 3.1 are not satisfied in this example because* $\inf_{a \in \boldsymbol{A}^G} q_0(a) = 0$.

We can strengthen the assertions of Corollary 3.1 if we impose the following set of compactness-continuity conditions and nonnegativity conditions.

Condition 3.2 *(a)* $\boldsymbol{A}^G$ *and* $\boldsymbol{A}^I$ *are compact.*

(b) The functions $\{c_i^G\}_{i=0}^J$ *and* $\{c_i^I\}_{i=0}^J$ *are* $[0,\infty]$*-valued and lower semicontinuous on* $\boldsymbol{X} \times \boldsymbol{A}^G$ *and* $\boldsymbol{X} \times \boldsymbol{A}^I$, *respectively.*

(c) For each bounded continuous function f *on* $\boldsymbol{X}$, *the functions* $(x,a) \in \boldsymbol{X} \times \boldsymbol{A}^G \to \int_{\boldsymbol{X}} f(y)\tilde{q}(dy|x,a)$ *and* $(x,b) \in \boldsymbol{X} \times \boldsymbol{A}^I \to \int_{\boldsymbol{X}} f(y)Q(dy|x,b)$ *are continuous.*

The next statement is the main solvability result concerning the gradual-impulsive control problem (3), obtained by an application of the proposed method (see Remark 3.1) for studying problem (3).

Theorem 3.3 *Suppose Conditions 3.1 and 3.2 are satisfied, and there is some* $\epsilon > 0$ *such that* $q_x(a) \geq \epsilon > 0$ *for all* $x \in \boldsymbol{X}$ *and* $a \in \boldsymbol{A}^G$. *If there exists a feasible policy* $\sigma' \in \Sigma$ *with a finite value, i.e., it satisfies the constraints in the gradual-impulsive control problem (3) and verifies* $\hat{W}_0(x_0,\sigma') < \infty$, *then there exists an optimal policy* $\sigma = \{\sigma^{(0)}_n, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ *in such a form that*

$$\begin{aligned}
&\sigma^{(0)}_n(\{\infty\} \times d\hat{b}|\hat{h}_n) = \sigma^{(0)}(\{\infty\} \times d\hat{b}|x_n) = \mu(x_n)\bar{\varphi}(d\hat{b}|x_n),\\
&\sigma^{(0)}_n(\{0\} \times d\hat{b}|x_n) = (1-\mu(x_n))\bar{\varphi}(d\hat{b}|x_n),\\
&\hat{F}_n(\hat{h}_n)_t(da) \equiv \hat{F}(x_n)(da),
\end{aligned} \tag{7}$$

where μ and $\hat{F}$ are $[0,1]$- and $\mathcal{R}(\boldsymbol{A}^G)$-valued measurable mappings on $\boldsymbol{X}$ and $\overline{\varphi}$ is a stochastic kernel. (A policy $\sigma = \{\sigma_n^{(0)}, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ in the form of (7) is called stationary.)

We finish this section with a few words on the role of Condition 3.2. According to Theorems 3.1 and 3.2, the gradual-impulsive control problem can be reduced to a standard CTMDP with gradual control only. This result holds for cost rates in general signs. Moreover, the induced standard CTMDP problem can be further reduced to a discrete-time Markov decision process with total cost criteria, see e.g., [16, 19, 29]. Condition 3.2 in fact is a sufficient condition for the optimality results (such as the existence of a stationary optimal control policy) for the induced discrete-time Markov decision process. The situation becomes more complicated when the cost rates are general signed, as the induced discrete-time problem will be with general signed cost functions. This case is quite challenging to handle and admits pathological scenarios unless further restrictions (stability-type) are imposed on the controlled process. We refer the interested readers to [1, 9, 10, 20] for further details.

4 Proofs of the main statements

4.1 General CTMDP model and known facts

To serve the proofs of Theorems 3.1, 3.2 and 3.3, we will make use of the following more general CTMDP model, introduced in [29]. Compared to the standard CTMDP model, it allows a richer class of control strategies. In fact, the standard CTMDP model can be viewed as a submodel or induced model of the general CTMDP model, to be described next.

The system primitives of the general CTMDP model are the same as those in the standard CTMDP model: $\mathcal{M}^{GO} := \{\mathbf{X}, \mathbf{A}, q^{GO}, \{c_i^{GO}\}_{i=0}^{J}\}$.

In the general CTMDP model $\mathcal{M}^{GO}$, the implementation of a strategy is in two steps. Firstly, the decision maker selects a Borel space $\mathbf{\Xi}$ upfront before the process starts [2] Secondly, once the process starts, a decision epoch occurs after each natural jump of the controlled process (except for the initial decision epoch at time zero). At each decision epoch, one selects $\xi \in \mathbf{\Xi}$, and based on ξ and other past information, one selects the relaxed control $\rho \in \mathcal{R}(\mathbf{A})$ until the next decision epoch occurs. We sketch the more rigorous construction as follows.

Suppose the Borel space $\mathbf{\Xi}$ was selected by the decision maker. Then it induces the corresponding sample space Ω as the union of $(\mathbf{\Xi} \times (\mathbf{X} \times \mathbf{\Xi} \times (0,\infty))^{\infty})$ and the collection of sequences in the form $(\xi_0, x_0, \xi_1, \theta_1, x_1, \xi_2, \dots, \theta_{m-1}, x_{m-1}, \xi_m, \theta_m, x_\infty, \xi_\infty, \infty, x_\infty, \xi_\infty, \dots)$, where $m \geq 1$, and $x_\infty \notin \mathbf{X}$, $\xi_\infty \notin \mathbf{\Xi}$ are isolated points. We endow Ω with the σ-algebra $\mathcal{F}$ obtained as the trace of $\mathcal{B}(\mathbf{\Xi}_\infty \times (\mathbf{X}_\infty \times \mathbf{\Xi}_\infty \times (0,\infty])^{\infty})$ on Ω, where $\mathbf{X}_\infty = \mathbf{X} \bigcup \{x_\infty\}$ and $\mathbf{\Xi}_\infty := \mathbf{\Xi} \bigcup \{\xi_\infty\}$.

For each $\omega \in \Omega$, define $\theta_0 := 0$, $t_n := \sum_{i=0}^{n} \theta_i$, $h_n := (\xi_0, x_0, \xi_1, \theta_1, x_1, \dots, \xi_n, \theta_n, x_n)$ for each $n \geq 0$, with $h_0 := (\xi_0, x_0)$. The collection of all possible h_n is denoted as $\mathbf{H}_n$ for each $n \geq 0$. Let us put $t_\infty := \lim_{n\to\infty} t_n$, which exists. When regarded as coordinate variables, we use capital letters Θ_n, T_n, X_n, Ξ_n and H_n corresponding to $\theta_n, t_n, x_n, \xi_n$ and h_n. The state process $\{X(t)\}_{t\geq 0}$ is defined by $X(t) := X_n$ if $T_n \leq t < T_{n+1}$ for some $n \geq 0$, and $X(t) := x_\infty$ if $t \geq T_\infty$. As usual, we omit ω whenever the context excludes confusion.

Remark 4.1 *Note that the objects such as Ω, $\boldsymbol{H}_n$ depend on $\mathbf{\Xi}$, but we do not indicate that dependence for brevity. Also we beg the reader's pardon for using the common notations such as Ω, $\boldsymbol{H}_n$ and t_n, x_n, etc both in the standard CTMDP model and in the general CTMDP model. The reasons are double-folded: first, the context will always clarify the underlying model being concerned with; and second, the*

[2] The additional flexibility in selecting the Borel space $\mathbf{\Xi}$ was introduced into the model of CTMDPs in [29]. A suitable selection of $\mathbf{\Xi}$ can lead to much convenience and useful consequences. For example, in Definition 7 of [29], $\mathbf{\Xi}$ was taken to be the countable product $((-\infty,\infty) \times \mathbf{A})^{\infty}$, which virtually introduced some artificial Poissonian jumps into the model. In relation to the current paper, $\mathbf{\Xi}$ will be selected to be $[0,\infty] \times \mathbf{A}^I$ in the proof of Theorem 3.2 below.

standard CTMDP model can be viewed as a submodel of of the general CTMDP model, as explained below, see the paragraph above Definition 4.2.

Definition 4.1 *A strategy S in the general CTMDP model $\mathcal{M}^{GO}$ is the following object: $S = \{\mathbf{\Xi}, \{\zeta_n\}_{n=0}^{\infty}, \{F_n\}_{n=0}^{\infty}\}$, where $\mathbf{\Xi}$ is a Borel space; $\zeta_0 \in \mathcal{P}(\mathbf{\Xi})$; for each $n \geq 1$, $\zeta_n(d\xi|h_{n-1})$ is a stochastic kernel on $\mathcal{B}(\mathbf{\Xi})$ given $h_{n-1} \in \boldsymbol{H}_{n-1}$; for each $n \geq 0$, $F_n(h_n, \xi_{n+1})$ is a measurable mapping on $\boldsymbol{H}_n \times \mathbf{\Xi}$ taking values in $\mathcal{R}(\boldsymbol{A})$. Here, in line with the previous descriptions, the first element $\mathbf{\Xi}$ of S is the Borel space selected upfront under the strategy S, and then the notations such as $\boldsymbol{H}_n$ are understood accordingly.*

Given a strategy $S = \{\mathbf{\Xi}, \{\zeta_n\}_{n=0}^{\infty}, \{F_n\}_{n=0}^{\infty}\}$ and initial state $x_0 \in \mathbf{X}$, there is a unique probability measure $\mathrm{P}_{x_0}^S$ on $(\Omega, \mathcal{F})$ such that $\mathrm{P}_{x_0}^S(\Xi_0 \in d\xi, X_0 \in dx) = \zeta_0(d\xi)\delta_{x_0}(dx)$, and for each $n \geq 1$ and $\Gamma_0 \in \mathcal{B}(\mathbf{\Xi})$, $\Gamma_1 \in \mathcal{B}([0,\infty))$, $\Gamma_2 \in \mathcal{B}(\mathbf{X})$,

$$\begin{aligned}
&\mathrm{P}_{x_0}^S(\Xi_n \in \Gamma_0,\ \Theta_n \in \Gamma_1,\ X_n \in \Gamma_2|H_{n-1}) \\
= &\int_{\Gamma_0}\int_{\Gamma_1} e^{-\int_0^s q_{X_{n-1}}^{GO}(F_{n-1}(H_{n-1},\xi)_t)dt}\tilde{q}^{GO}(\Gamma_2|X_{n-1}, F_{n-1}(H_{n-1},\xi)_s)ds\zeta_n(d\xi|H_{n-1}); \\
&\mathrm{P}_{x_0}^S(\Xi_n \in \Gamma_0,\ \Theta_n = \infty,\ X_n = x_\infty|H_{n-1}) = \int_{\Gamma_0} e^{-\int_0^\infty q_{X_{n-1}}^{GO}(F_{n-1}(H_{n-1},\xi)_t)dt}\zeta_n(d\xi|H_{n-1});
\end{aligned}$$

and

$$\mathrm{P}_{x_0}^S(\Xi_n \in \Gamma_0,\ \Theta_n = \infty,\ X_n \in \Gamma_2|H_{n-1}) = \mathrm{P}_{x_0}^S(\Xi_n \in \Gamma_0,\ \Theta_n \in \Gamma_1,\ X_n = x_\infty|H_{n-1}) = 0.$$

(Recall Remark 2.2.) Let the expectation corresponding to $\mathrm{P}_{x_0}^S$ be denoted as $\mathrm{E}_{s_0}^S$. We put

$$\begin{aligned}
W_i(x_0,S) &:= \mathrm{E}_{x_0}^S\left[\sum_{n=0}^{\infty} I\{T_n < \infty\}\int_{T_n}^{T_{n+1}} c_i^{GO}(X_n, F_n(H_n,\Xi_{n+1})_{t-T_n})dt\right] \\
&:= \mathrm{E}_{x_0}^S\left[\sum_{n=0}^{\infty} I\{T_n < \infty\}\int_{T_n}^{T_{n+1}} {c^{GO}}_i^+(X_n, F_n(H_n,\Xi_{n+1})_{t-T_n})dt\right] \\
&\quad -\mathrm{E}_{x_0}^S\left[\sum_{n=0}^{\infty} I\{T_n < \infty\}\int_{T_n}^{T_{n+1}} {c^{GO}}_i^-(X_n, F_n(H_n,\Xi_{n+1})_{t-T_n})dt\right],
\end{aligned}$$

with $\infty - \infty := \infty$ being accepted here. We may also write

$$\begin{aligned}
W_i(x_0,S) &= \mathrm{E}_{x_0}^S\left[\sum_{n=0}^{\infty}\int_0^{\Theta_{n+1}} {c^{GO}}_i^+(X_n, F_n(H_n,\Xi_{n+1})_t)dt\right] \\
&\quad -\mathrm{E}_{x_0}^S\left[\sum_{n=0}^{\infty}\int_0^{\Theta_{n+1}} {c^{GO}}_i^-(X_n, F_n(H_n,\Xi_{n+1})_t)dt\right].
\end{aligned}$$

We consider two important subclasses of strategies in the general CTMDP model $\mathcal{M}^{GO}$, which will be referred to in the next subsection. In case we only consider the class of strategies $S = \{\mathbf{\Xi}, \{\zeta_n\}_{n=0}^{\infty}, \{F_n\}_{n=0}^{\infty}\}$, where $\mathbf{\Xi}$ is a singleton say $\{\xi'\}$, then we can and will omit the ξ terms from ω and F_n, and retrieve the standard CTMDP model described in Section 2. A strategy $\overline{S} = \{\overline{F}_n\}_{n=0}^{\infty}$ in the standard CTMDP model clearly identifies a strategy (with the first element being a singleton) in the general CTMDP model. Another subclass of strategies of interest is the following one.

Definition 4.2 *In the general CTMDP model, a strategy* $S = \{\boldsymbol{A}, \{\zeta_n\}_{n=0}^{\infty}, \{F_n\}_{n=0}^{\infty}\}$, *where, for each* $n \geq 0$, $F_n(h_n, \xi)_t(da) \equiv \delta_\xi(da)$, *is called standard randomized. A standard randomized strategy* S *in the general CTMDP model is identified with* $\{\zeta_n\}_{n=0}^{\infty}$. *If for each* $n \geq 1$, $\zeta_n(d\xi|h_{n-1})$ *depends on* h_{n-1} *only through* x_{n-1}, *we write* $\zeta_n(d\xi|h_{n-1}) = \zeta_n(d\xi|x_{n-1})$, *and* ζ_0 *can be discarded, in which case, we call the strategy standard Markov randomized, and identify it with* $\zeta = \{\zeta_n\}_{n=1}^{\infty}$. *If, additionally,* $\zeta_n(da|x) = \zeta^s(da|x)$ *for each* $n \geq 1$ *for some stochastic kernel* ζ^s, *then the strategy is called standard stationary, and is identified as* ζ^s.

A standard (Markov, stationary) randomized strategy is also called a (Markov, stationary) standard ξ-strategy in [29]. The submodel induced by concentrating only on the class of standard randomized strategies is termed by Feinberg [14] an ESMDP (exponential semi-Markov decision process) model. For instance, under a standard Markov randomized strategy ζ, the process $\{X(t)\}_{t\geq 0}$ is a semi-Markov process.

Theorems 3.1 and 3.2 involve the comparison of the performance measures $\{\hat{W}_i(x_0, \sigma)\}_{i=0}^{J}$ and $\{W_i(x_0, \overline{S})\}_{i=0}^{J}$ in the gradual-impulsive control model and in the standard CTMDP model. More generally, the performance $W_i(x_0, S)$ of a strategy S in the general CTMDP model can be expressed as integrals with respect to detailed occupation measures, defined as follows.

Definition 4.3 *The detailed occupation measure* $\{\eta_n^S\}_{n=0}^{\infty}$ *of a strategy* $S = \{\boldsymbol{\Xi}, \{\zeta_n\}_{n=0}^{\infty}, \{F_n\}_{n=0}^{\infty}\}$ *in the general CTMDP model* $\mathcal{M}^{GO}$ *is the sequence of measures* η_n^S *on* $\mathcal{B}(\boldsymbol{X} \times \boldsymbol{A})$ *defined by*

$$\begin{aligned} \eta_n^S(dx \times da) &:= \mathrm{E}_{x_0}^S \left[I\{T_n < \infty\} \int_{T_n}^{T_{n+1}} I\{X_n \in dx\} F_n(H_n, \Xi_{n+1})_{t-T_n}(da) dt \right] \\ &= \mathrm{E}_{x_0}^S \left[\int_0^{\Theta_{n+1}} I\{X_n \in dx\} F_n(H_n, \Xi_{n+1})_t(da) dt \right], \end{aligned}$$

where the second equality holds automatically.

We will make use of the following fact quoted from Theorems 1 and 2 of [29] as well as their proofs therein, see also Feinberg [14, 16].

Proposition 4.1 *The following assertions hold.*

(a) Consider the general CTMDP model $\mathcal{M}^{GO}$. *For each strategy* $S = \{\boldsymbol{\Xi}, \{\zeta_n\}_{n=0}^{\infty}, \{F_n\}_{n=0}^{\infty}\}$, *there is a strategy* $\bar{S} = \{\bar{F}_n\}_{n=0}^{\infty}$ *in the standard CTMDP model* $\mathcal{M}^{GO}$ *such that* $\eta_n^S = \eta_n^{\bar{S}}$ *for each* $n \geq 0$.

(b) Assume that there is some $\epsilon > 0$ *satisfying* $q_x^{GO}(a) \geq \epsilon > 0$ *for each* $(x, a) \in \boldsymbol{X} \times \boldsymbol{A}$. *Then for each strategy* $\overline{S} = \{\overline{F}_n\}_{n=0}^{\infty}$ *in the standard CTMDP model* $\mathcal{M}^{GO}$, *there is a standard Markov randomized strategy* $\zeta = \{\zeta_n\}_{n=1}^{\infty}$ *in the general CTMDP model such that*

$$\eta_n^{\overline{S}}(dx \times da) = \eta_n^{\zeta}(dx \times da), \quad \int_{\boldsymbol{A}} q_x^{GO}(a) \eta_n^{\overline{S}}(dx \times da) = \mathrm{P}_{x_0}^{\zeta}(X_n \in dx) \tag{8}$$

for each $n \geq 0$. *Moreover, for each* $n \geq 0$, *one can take* ζ_{n+1} *as the stochastic kernel satisfying*

$$\zeta_{n+1}(da|x) \mathrm{P}_{x_0}^{\zeta}(X_n \in dx) = q_x^{GO}(a) \eta_n^{\overline{S}}(dx \times da). \tag{9}$$

Finally, if $\overline{S} = \{\overline{F}_n\}_{n=0}^{\infty}$ *is stationary, i.e., it is in such a form that* $\overline{F}_n(h_n)_t(da) \equiv \overline{F}(x_n)(da)$, *then one can take*

$$\zeta_{n+1}(da|x) \equiv \zeta^s(da|x) = \frac{q_x^{GO}(a) \overline{F}(x)(da)}{\int_{\boldsymbol{A}} q_x^{GO}(a) \overline{F}(x)(da)} \tag{10}$$

on $\mathcal{B}(\boldsymbol{A})$ *for each* $x \in \boldsymbol{X}$. *(Such a stationary strategy in the standard CTMDP model will be denoted as* $\overline{S} = \{\overline{F}\}$.*)*

4.2 Proofs of Theorems 3.1, 3.2 and 3.3

Proof of Theorem 3.1. Let $\overline{S} = \{\overline{F}_n\}_{n=0}^{\infty}$ be a fixed strategy in the standard CTMDP model. We will show that for some policy $\sigma = \{\sigma_n^{(0)}, \hat{F}_n\}_{n=0}^{\infty} \in \Sigma$ for the gradual-impulsive CTMDP model,

$$\hat{\mathrm{E}}_{x_0}^{\sigma}\left[l_i(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1})\right] = \mathrm{E}_{x_0}^{\overline{S}}\left[\int_0^{\Theta_{n+1}} c_i^{GO}(X_n, \overline{F}_n(H_n)_t)dt\right] \tag{11}$$

for each $n \geq 0$ and $i = 0, 1, \ldots, J$. It is without loss of generality to assume c_i^{GO} is $[0, \infty]$-valued, for otherwise, one would apply the reasoning below to c_i^{GO+} and c_i^{GO-}, separately.

Consider the standard Markov strategy $\zeta = \{\zeta_n\}_{n=1}^{\infty}$ in Proposition 4.1(b). Then

$$\begin{aligned}
&\mathrm{E}_{x_0}^{\overline{S}}\left[\int_0^{\Theta_{n+1}} c_i^{GO}(X_n, \overline{F}_n(H_n)_t)dt\right] = \int_{\mathbf{X}\times\mathbf{A}} c_i^{GO}(x,a)\eta_n^{\overline{S}}(dx\times da) = \int_{\mathbf{X}\times\mathbf{A}} c_i^{GO}(x,a)\eta_n^{\zeta}(dx\times da)\\
= &\int_{\mathbf{X}\times\mathbf{A}} \frac{c_i^{GO}(x,a)}{q_x^{GO}(a)} q_x^{GO}(a)\eta_n^{\zeta}(dx\times da) = \int_{\mathbf{X}\times\mathbf{A}} \frac{c_i^{GO}(x,a)}{q_x^{GO}(a)}\zeta_{n+1}(da|x)\mathrm{P}_{x_0}^{\zeta}(X_n\in dx)\\
= &\int_{\mathbf{X}\times\mathbf{A}^G} \frac{c_i^{GO}(x,a)}{q_x^{GO}(a)}\zeta_{n+1}(da|x)\mathrm{P}_{x_0}^{\zeta}(X_n\in dx) + \int_{\mathbf{X}\times\mathbf{A}^I} \frac{c_i^{GO}(x,a)}{q_x^{GO}(a)}\zeta_{n+1}(da|x)\mathrm{P}_{x_0}^{\zeta}(X_n\in dx),
\end{aligned} \tag{12}$$

where the second equality is by (8) and the forth equality is by (9). Let us define for each $n \geq 0$ a stochastic kernel $\tilde{\varphi}_n$ on $\mathcal{B}(\mathbf{A}^G)$ given $x \in \mathbf{X}$ by

$$\tilde{\varphi}_n(da|x) := \frac{\zeta_{n+1}(da \bigcap \mathbf{A}^G|x)}{\zeta_{n+1}(\mathbf{A}^G|x)} \tag{13}$$

for each $x \in \mathbf{X}$ where $\zeta_{n+1}(\mathbf{A}^G|x) > 0$; for all $x \in \mathbf{X}$ where $\zeta_{n+1}(\mathbf{A}^G|x) = 0$, we put $\tilde{\varphi}_n(da|x)$ as a fixed probability measure on $\mathbf{A}^G$.

Similarly, we define for each $n \geq 0$ a stochastic kernel $\bar{\varphi}_n$ on $\mathcal{B}(\mathbf{A}^I)$ given $x \in \mathbf{X}$ by

$$\bar{\varphi}_n(da|x) := \frac{\zeta_{n+1}(da \bigcap \mathbf{A}^I|x)}{\zeta_{n+1}(\mathbf{A}^I|x)}, \tag{14}$$

for each $x \in \mathbf{X}$ where $\zeta_{n+1}(\mathbf{A}^I|x) > 0$; for all $x \in \mathbf{X}$ where $\zeta_{n+1}(\mathbf{A}^I|x) = 0$, we put $\bar{\varphi}_n(da|x)$ as a fixed probability measure on $\mathbf{A}^I$.

Now we continue from (12):

$$\begin{aligned}
\mathrm{E}_{x_0}^{\overline{S}}\left[\int_0^{\Theta_{n+1}} c_i^{GO}(X_n, \overline{F}_n(H_n)_t)dt\right] &= \int_{\mathbf{X}}\int_{\mathbf{A}^G} \frac{c_i^G(x,a)}{q_x(a)}\tilde{\varphi}_n(da|x)\zeta_{n+1}(\mathbf{A}^G|x)\mathrm{P}_{x_0}^{\zeta}(X_n\in dx)\\
&\quad + \int_{\mathbf{X}}\int_{\mathbf{A}^I} c_i^I(x,a)\bar{\varphi}_n(da|x)\zeta_{n+1}(\mathbf{A}^I|x)\mathrm{P}_{x_0}^{\zeta}(X_n\in dx).
\end{aligned} \tag{15}$$

Let us further define for each $n \geq 0$ a stochastic kernel $\hat{F}_n(x)(da)$ on $\mathcal{B}(\mathbf{A}^G)$ given $x \in \mathbf{X}$ by

$$\hat{F}_n(x)(da) := \frac{\frac{1}{q_x(a)}\tilde{\varphi}_n(da|x)}{\int_{\mathbf{A}^G}\frac{1}{q_x(a)}\tilde{\varphi}_n(da|x)}, \ \forall\ x \in \mathbf{X}. \tag{16}$$

Then

$$\frac{\int_{\mathbf{A}^G} c_i^G(x,a)\hat{F}_n(x)(da)}{\int_{\mathbf{A}^G} q_x^{GO}(a)\hat{F}_n(x)(da)} = \frac{\int_{\mathbf{A}^G} c_i^G(x,a)\hat{F}_n(x)(da)}{\int_{\mathbf{A}^G} q_x(a)\hat{F}_n(x)(da)} = \int_{\mathbf{A}^G}\frac{c_i^G(x,a)}{q_x(a)}\tilde{\varphi}_n(da|x),$$

and so from (15)

$$\begin{aligned}
&\mathrm{E}^{\overline{S}}_{x_0}\left[\int_0^{\Theta_{n+1}}\int_{\mathbf{A}} c_i^{GO}(X_n,a)\overline{F}_n(H_n)_t(da)dt\right]\\
=&\int_{\mathbf{X}}\frac{\int_{\mathbf{A}^G} c_i^G(x,a)\hat{F}_n(x)(da)}{\int_{\mathbf{A}^G} q_x^{GO}(a)\hat{F}_n(x)(da)}\zeta_{n+1}(\mathbf{A}^G|x)\mathrm{P}^{\zeta}_{x_0}(X_n\in dx)\\
&+\int_{\mathbf{X}}\int_{\mathbf{A}^I} c_i^I(x,a)\bar{\varphi}_n(da|x)\zeta_{n+1}(\mathbf{A}^I|x)\mathrm{P}^{\zeta}_{x_0}(X_n\in dx). \qquad (17)
\end{aligned}$$

Now consider the policy $\sigma=\{\sigma_n^{(0)},\hat{F}_n\}_{n=0}^{\infty}$ in the gradual-impulsive control model defined by $\hat{F}_n(x_n)_t(da)$ introduced above and

$$\sigma_n^{(0)}(\{\infty\}\times d\hat{b}|x_n)=\zeta_{n+1}(\mathbf{A}^G|x_n)\bar{\varphi}_n(d\hat{b}|x_n),\ \sigma_n^{(0)}(\{0\}\times d\hat{b}|x_n)=(1-\zeta_{n+1}(\mathbf{A}^G|x_n))\bar{\varphi}_n(d\hat{b}|x_n), \quad (18)$$

In particular, $\sigma_n^{(0)}(\{\infty\}\bigcup\{0\}\times\mathbf{A}^I|x_n)=1$.

Note that on $\{\hat{\Theta}_n<\infty\}$,

$$\begin{aligned}
&\hat{\mathrm{E}}^{\sigma}_{x_0}\left[l_i(\hat{X}_n,\hat{A}_n,\hat{X}_{n+1})|\hat{H}_n\right]\\
=&\hat{\mathrm{E}}^{\sigma}_{x_0}\left[I\{\hat{\Theta}_{n+1}<\hat{C}_n\}\int_0^{\hat{\Theta}_{n+1}}\int_{\mathbf{A}^G} c_i^G(X_n,a)\hat{F}_n(X_n)(da)dt|\hat{H}_n\right]\\
&+\hat{\mathrm{E}}^{\sigma}_{x_0}\left[I\{\hat{\Theta}_{n+1}=\hat{C}_n\}c_i^I(X_n,\hat{B}_n)|\hat{H}_n\right]\\
=&\zeta_{n+1}(\mathbf{A}^G|X_n)\frac{\int_{\mathbf{A}^G} c_i^G(X_n,a)\hat{F}_n(X_n)(da)}{\int_{\mathbf{A}^G} q_{X_n}^{GO}(a)\hat{F}_n(X_n)(da)}+\zeta_{n+1}(\mathbf{A}^I|X_n)\int_{\mathbf{A}^I} c_i^I(X_n,a)\bar{\varphi}_n(da|X_n)\\
=&\hat{\mathrm{E}}^{\sigma}_{x_0}\left[l_i(\hat{X}_n,\hat{A}_n,\hat{X}_{n+1})|X_n\right],
\end{aligned}$$

where the second equality is by (18). Comparing this (in particular, the expression in the second to the last line) with (17), we see that, for (11) and thus for the statement of this theorem, it remains to show that $\mathrm{P}^{\zeta}_{x_0}(X_n\in dx)=\hat{\mathrm{P}}^{\sigma}_{x_0}(X_n\in dx)$ as follows. This relation automatically holds when $n=0$, with both sides of the equality being $\delta_{x_0}(dx)$. Assume for induction that it also holds for the case of n. Then

$$\begin{aligned}
&\hat{\mathrm{P}}^{\sigma}_{x_0}(X_{n+1}\in dx)=\hat{\mathrm{E}}^{\sigma}_{x_0}\left[\hat{\mathrm{P}}^{\sigma}_{x_0}(X_{n+1}\in dx|\hat{H}_n,\hat{C}_n,\hat{B}_n)(I\{\hat{C}_n=\infty\}+I\{\hat{C}_n=0\})\right]\\
=&\hat{\mathrm{E}}^{\sigma}_{x_0}\left[\int_0^{\infty}\int_{\mathbf{A}^G}\tilde{q}(dx|X_n,a)\hat{F}_n(X_n)(da)e^{-\int_{\mathbf{A}^G} q_{X_n}(a)\hat{F}_n(X_n)(da)t}dt\zeta_{n+1}(\mathbf{A}^G|X_n)\right.\\
&\left.+\int_{\mathbf{A}^I} Q(dx|X_n,a)\bar{\varphi}_n(da|X_n)\zeta_{n+1}(\mathbf{A}^I|X_n)\right]\\
=&\hat{\mathrm{E}}^{\sigma}_{x_0}\left[\frac{\int_{\mathbf{A}^G}\tilde{q}(dx|X_n,a)\hat{F}_n(X_n)(da)}{\int_{\mathbf{A}^G} q_{X_n}(a)\hat{F}_n(X_n)(da)}\zeta_{n+1}(\mathbf{A}^G|X_n)+\int_{\mathbf{A}^I} Q(dx|X_n,a)\bar{\varphi}_n(da|X_n)\zeta_{n+1}(\mathbf{A}^I|X_n)\right]\\
=&\hat{\mathrm{E}}^{\sigma}_{x_0}\left[\int_{\mathbf{A}^G}\frac{\tilde{q}^{GO}(dx|X_n,a)}{q_{X_n}^{GO}(a)}\tilde{\varphi}_n(da|X_n)\zeta_{n+1}(\mathbf{A}^G|X_n)+\int_{\mathbf{A}^I}\frac{\tilde{q}^{GO}(dx|X_n,a)}{q_{X_n}^{GO}(a)}\bar{\varphi}_n(da|X_n)\zeta_{n+1}(\mathbf{A}^I|X_n)\right]\\
=&\hat{\mathrm{E}}^{\sigma}_{x_0}\left[\int_{\mathbf{A}}\frac{\tilde{q}^{GO}(dx|X_n,a)}{q_{X_n}^{GO}(a)}\zeta_{n+1}(da|X_n)\right]=\mathrm{E}^{\zeta}_{x_0}\left[\int_{\mathbf{A}}\frac{\tilde{q}^{GO}(dx|X_n,a)}{q_{X_n}^{GO}(a)}\zeta_{n+1}(da|X_n)\right]=\mathrm{P}^{\zeta}_{x_0}(X_{n+1}\in dx),
\end{aligned}$$

where the second equality is by (18), the forth equality is by (16), the third to the last equality is by the definitions of $\tilde{\varphi}_n$, $\bar{\varphi}_n$, and the second to the last equality is by the inductive supposition, and the

last equality holds because ζ is a Markov strategy in the ESMDP model. Therefore, $\mathrm{P}^{\overline{S}}_{x_0}(X_n \in dx) = \hat{\mathrm{P}}^{\sigma}_{x_0}(X_n \in dx)$ for all $n \geq 0$, as required. The first assertion of this statement is thus proved.

The last assertion of this statement now follows from the above proof, (16) and (18). □

Proof of Theorem 3.2. Note that in the DTMDP model corresponding to the gradual-impulsive CTMDP model, $l_i(\hat{x}, \hat{a}, \hat{y})$ and $p(d\hat{y}|\hat{x}, \hat{a})$ depend on $\hat{x} = (\hat{\theta}, x)$ only through their second coordinate. Therefore, for the gradual-impulsive control problem (3) it suffices to concentrate on the class of policies $\sigma = \{\sigma^{(0)}_n, \hat{F}_n\}^{\infty}_{n=0}$, where $\sigma^{(0)}_n(d\hat{c} \times d\hat{b}|\hat{h}_n)$ and $\hat{F}_n(\hat{h}_n, \hat{c}_n, \hat{b}_n)_t(da)$ do not depend on the inessential information $\{\hat{\theta}_m\}^n_{m=1}$, see [15]. We will accordingly write

$$\begin{aligned}\sigma^{(0)}_n(d\hat{c} \times d\hat{b}|\hat{h}_n) &= \sigma^{(0)}_n(d\hat{c} \times d\hat{b}|x_0, \hat{c}_0, \hat{b}_0, x_1, \hat{c}_1, \hat{b}_1, \ldots, x_n, \hat{c}_n, \hat{b}_n),\\ \hat{F}_n(\hat{h}_n, \hat{c}_n, \hat{b}_n)_t(da) &= \hat{F}_n(x_0, \hat{c}_0, \hat{b}_0, x_1, \hat{c}_1, \hat{b}_1, \ldots, x_n, \hat{c}_n, \hat{b}_n)_t(da).\end{aligned}$$

Let such a policy $\sigma = \{\sigma^{(0)}_n, \hat{F}_n\}^{\infty}_{n=0} \in \Sigma$ be fixed. For the statement of this theorem, we will construct a strategy $\overline{S} = \{\overline{F}_n\}^{\infty}_{n=0}$ in the standard CTMDP model $\mathcal{M}^{GO}$ such that (11) holds for all $n \geq 0$. As in the proof of Theorem 3.1, we may assume that c^{GO}_i is $[0, \infty]$-valued. Moreover, we will further assume in this proof that c^{GO}_i is bounded. This is without loss of generality, because one can deduce the general case by applying the claimed relation to $\min\{c^{GO}_i, N\}$ and pass to the limit as $N \to \infty$ with the help of monotone convergence theorem.

First, let us consider the general CTMDP model $\mathcal{M}^{GO} = \{\mathbf{X}, \mathbf{A}, q^{GO}, \{c^{GO}_i\}^J_{i=0}\}$, and define a strategy $S = \{\mathbf{\Xi}, \{\zeta_n\}^{\infty}_{n=0}, \{F_n\}^{\infty}_{n=0}\}$ as follows:

$$\mathbf{\Xi} := [0, \infty] \times \mathbf{A}^I$$

and for $n \geq 0$, $\xi_1 = (\hat{c}_0, \hat{b}_0), \ldots, \xi_{n+1} = (\hat{c}_n, \hat{b}_n)$,

$$\begin{aligned}&\zeta_{n+1}(dc \times db|x_0, \xi_1, x_1, \xi_2, \ldots, \xi_n, x_n) = \zeta_{n+1}(dc \times db|x_0, (\hat{c}_0, \hat{b}_0), x_1, (\hat{c}_1, \hat{b}_1), \ldots, x_n)\\ :=\ &\sigma^{(0)}_n(dc \times db|x_0, \hat{c}_0, \hat{b}_0, x_1, \hat{c}_1, \hat{b}_1, \ldots, x_n);\end{aligned} \tag{19}$$

and

$$\begin{aligned}&F_n((x_0, \xi_1, x_1, \xi_2, \ldots, x_n), \xi_{n+1})_t(da) = F_n((x_0, (\hat{c}_0, \hat{b}_0), x_1, (\hat{c}_1, \hat{b}_1), \ldots, x_n), (\hat{c}_n, \hat{b}_n))_t(da)\\ :=\ &\begin{cases}\hat{F}_n(x_0, \hat{c}_0, \hat{b}_0, x_1, \hat{c}_1, \hat{b}_1, \ldots, x_n, \hat{c}_n, \hat{b}_n)_t(da) & \text{if } t < \hat{c}_n;\\ \delta_{\hat{b}_n}(da) & \text{if } t \geq \hat{c}_n,\end{cases}\end{aligned} \tag{20}$$

where in the right hand side, we do not indicate the inessential argument of F_n as done for $\hat{F}_n$, and in accordance with Definition 4.1, $(\hat{c}_0, \hat{b}_0) \in \mathbf{\Xi}, \ldots, (\hat{c}_n, \hat{b}_n) \in \mathbf{\Xi}$ correspond to $(c_1, b_1) \in \mathbf{\Xi}, \ldots, (c_{n+1}, b_{n+1}) \in \mathbf{\Xi}$, and there is no dependence on ξ_0.

Let us justify the following equality:

$$\hat{\mathrm{E}}^{\sigma}_{x_0}\left[l_i(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1})\right] = \mathrm{E}^{S}_{x_0}\left[\int_0^{\Theta_{n+1}} c^{GO}_i(X_n, F_n(H_n)_t)dt\right] \tag{21}$$

for each $n \geq 0$ and $i = 0, 1, \ldots, J$.

It holds for each $n \geq 0$ that

$$\begin{aligned}&\mathrm{E}^{S}_{x_0}\left[\int_0^{\Theta_{n+1}} \int_{\mathbf{A}} c^{GO}_i(X_n, a) F_n((X_0, (C_1, B_1), X_1, \ldots, (C_n, B_n), X_n), (C_{n+1}, B_{n+1})_t(da)dt\right]\\ =\ &\mathrm{E}^{S}_{x_0}\left[\mathrm{E}^{S}_{x_0}\left[\int_0^{\Theta_{n+1}} \int_{\mathbf{A}} c^{GO}_i(X_n, a) F_n(H^-_n)_t(da)dt|H^-_n\right]\right],\end{aligned}$$

where

$$H_n^- := (X_0, (C_1, B_1) = \Xi_1, X_1, \ldots, (C_n, B_n) = \Xi_n, X_n, (C_{n+1}, B_{n+1}) = \Xi_{n+1}).$$

The conditional expectation in the previous equality can be written as

$$\begin{aligned}
& \mathrm{E}_{x_0}^S \left[\int_0^{\Theta_{n+1}} \int_{\mathbf{A}} c_i^{GO}(X_n, a) F_n(H_n^-)_t(da)dt | H_n^- \right] \\
= & \int_0^\infty \int_{\mathbf{A}} c_i^{GO}(X_n, a) F_n(H_n^-)_t(da) e^{-\int_0^t \int_{\mathbf{A}} q_{X_n}^{GO}(a) F_n(H_n^-)_s(da)ds} dt \\
= & \int_0^{C_{n+1}} \int_{\mathbf{A}^G} c_i^G(X_n, a) \hat{F}_n(H_n^-)_t(da) e^{-\int_0^t \int_{\mathbf{A}^G} q_{X_n}(a) \hat{F}_n(H_n^-)_s(da)ds} dt \\
& + I\{C_{n+1} < \infty\} \int_{C_{n+1}}^\infty c_i^I(X_n, B_{n+1}) e^{-(t - C_{n+1})} e^{-\int_0^{C_{n+1}} \int_{\mathbf{A}^G} q_{X_n}(a) \hat{F}_n(H_n^-)_s(da)ds} dt \\
= & \int_0^{C_{n+1}} \int_{\mathbf{A}^G} c_i^G(X_n, a) \hat{F}_n(H_n^-)_t(da) e^{-\int_0^t \int_{\mathbf{A}^G} q_{X_n}(a) \hat{F}_n(H_n^-)_s(da)ds} dt \\
& + I\{C_{n+1} < \infty\} c_i^I(X_n, B_{n+1}) e^{-\int_0^{C_{n+1}} \int_{\mathbf{A}^G} q_{X_n}(a) \hat{F}_n(H_n^-)_s(da)ds},
\end{aligned} \tag{22}$$

where the second equality is by (20) and (5).

On the other hand, with the notation

$$\hat{H}_n^- := (X_0, \hat{C}_0, \hat{B}_0, X_1, \hat{C}_1, \hat{B}_1, \ldots, X_n, \hat{C}_n, \hat{B}_n)$$

being introduced, we have

$$\hat{\mathrm{E}}_{x_0}^\sigma \left[l_i(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1}) \right] = \hat{\mathrm{E}}_{x_0}^\sigma \left[\hat{\mathrm{E}}_{x_0}^\sigma \left[l_i(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1}) | \hat{H}_n^- \right] \right],$$

where, according to (2),

$$\begin{aligned}
& \hat{\mathrm{E}}_{x_0}^\sigma \left[l_i(\hat{X}_n, \hat{A}_n, \hat{X}_{n+1}) | \hat{H}_n^- \right] \\
= & \hat{\mathrm{E}}_{x_0}^\sigma \left[\int_0^{\hat{\Theta}_{n+1}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_t) dt + I\{\hat{\Theta}_{n+1} = \hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n) | \hat{H}_n^- \right].
\end{aligned} \tag{23}$$

For each $m \geq 1$,

$$\begin{aligned}
& \hat{\mathrm{E}}_{x_0}^\sigma \left[\int_0^{\hat{\Theta}_{n+1}} e^{-\frac{t}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_t) dt + I\{\hat{\Theta}_{n+1} = \hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n) | \hat{H}_n^- \right] \\
= & \int_0^{\hat{C}_n} \int_0^t e^{-\frac{s}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_s) ds q_{X_n}(\hat{F}_n(\hat{H}_n^-)_t) e^{-\int_0^t q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s)ds} dt \\
& + \int_0^{\hat{C}_n} e^{-\frac{s}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_s) ds e^{-\int_0^{\hat{C}_n} q_{X_n}(\hat{F}_n(\hat{H}_n)_s)ds} \\
& + I\{\hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n) e^{-\int_0^{\hat{C}_n} q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s)ds},
\end{aligned} \tag{24}$$

where the equality is by (1) applied to the function

$$\begin{aligned}
& g(t, y) \equiv \int_0^t e^{-\frac{s}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_s) ds + I\{t = \hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n), \ \forall \ (t, y) \in [0, \infty) \times \mathbf{X}, \\
& g(\infty, x_\infty) \equiv \int_0^\infty e^{-\frac{s}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_s) ds,
\end{aligned}$$

which actually does not depend on the second component in its argument.

By integration by parts (recall that c_i^G is bounded nonnegative-valued as assumed without loss of generality in the beginning of this proof),

$$\begin{aligned}
&\int_0^{\hat{C}_n}\int_0^t e^{-\frac{s}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_s) ds q_{X_n}(\hat{F}_n(\hat{H}_n^-)_t) e^{-\int_0^t q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s) ds} dt \\
= &\int_0^{\hat{C}_n} e^{-\frac{t}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_t) e^{-\int_0^t q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s) ds} dt \\
&- \int_0^{\hat{C}_n} e^{-\frac{s}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_s) ds e^{-\int_0^{\hat{C}_n} q_{X_n}(\hat{F}_n(\hat{H}_n)_s) ds}.
\end{aligned}$$

Consequently, from (24), we see that

$$\begin{aligned}
&\hat{\mathrm{E}}_{x_0}^{\sigma}\left[\int_0^{\hat{\Theta}_{n+1}} e^{-\frac{t}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_t) dt + I\{\hat{\Theta}_{n+1} = \hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n) | \hat{H}_n^-\right] \\
= &\int_0^{\hat{C}_n} e^{-\frac{t}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_t) e^{-\int_0^t q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s) ds} dt + I\{\hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n) e^{-\int_0^{\hat{C}_n} q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s) ds}.
\end{aligned}$$

Passing to the limit on the both sides of this equality as $m \to \infty$ with the help of the monotone convergence theorem, we see from (23) that

$$\begin{aligned}
&\hat{\mathrm{E}}_{x_0}^{\sigma}\left[l_i(\hat{X}, \hat{A}_n, \hat{X}_{n+1}) | \hat{H}_n^-\right] \\
= &\lim_{m\to\infty} \hat{\mathrm{E}}_{x_0}^{\sigma}\left[\int_0^{\hat{\Theta}_{n+1}} e^{-\frac{t}{m}} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_t) dt + I\{\hat{\Theta}_{n+1} = \hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n) | \hat{H}_n^-\right] \\
= &\int_0^{\hat{C}_n} c_i^G(X_n, \hat{F}_n(\hat{H}_n^-)_t) e^{-\int_0^t q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s) ds} dt + I\{\hat{C}_n < \infty\} c_i^I(X_n, \hat{B}_n) e^{-\int_0^{\hat{C}_n} q_{X_n}(\hat{F}_n(\hat{H}_n^-)_s) ds}.
\end{aligned}$$

Comparing this equality with (22), we see that, for the claimed relation (21), it remains to show that the distribution of H_n^- under $\mathrm{P}_{x_0}^S$ (restricted on $(\mathbf{X} \times \mathbf{\Xi})^{n+1}$) coincides with the one of $\hat{H}_n^-$ under $\hat{\mathrm{P}}_{x_0}^{\sigma}$ (restricted on $(\mathbf{X} \times \mathbf{\Xi})^{n+1}$) for each $n \geq 0$. We verify this statement by induction as follows.

The case of $n = 0$ automatically holds, because of (19) and $\mathrm{P}_{x_0}^S(X_0 \in dx) = \hat{\mathrm{P}}_{x_0}^{\sigma}(X_0 \in dx) =$

$\delta_{x_0}(dx)$. Suppose the statement holds for the case of n. Then

$$\begin{aligned}
&\mathrm{P}^S_{x_0}(H^-_n \in dh, X_{n+1} \in dx,\ C_{n+2} \in dc,\ B_{n+2} \in db)\\
=\ &\mathrm{E}^S_{x_0}\left[\mathrm{P}^S_{x_0}(H^-_n \in dh, X_{n+1} \in dx,\ C_{n+2} \in dc,\ B_{n+2} \in db|H^-_n)\right]\\
=\ &\mathrm{E}^S_{x_0}\left[I\{H^-_n \in dh\}\zeta_{n+2}(dc \times db|H^-_n, x)\int_0^\infty \tilde{q}^{GO}(dx|X_n, F_n(H^-_n)_t)e^{-\int_0^t q^{GO}_{X_n}(F_n(H^-_n)_s)ds}dt\right]\\
=\ &\mathrm{E}^S_{x_0}\left[I\{H^-_n \in dh\}\zeta_{n+2}(dc \times db|H^-_n, x)\left\{\int_0^{C_{n+1}} \tilde{q}(dx|X_n, \hat{F}_n(H^-_n)_t)e^{-\int_0^t q_{X_n}(\hat{F}_n(H^-_n)_s)ds}dt\right.\right.\\
&\left.\left.+I\{C_{n+1} < \infty\}\int_{C_{n+1}}^\infty Q(dx|X_n, B_{n+1})e^{-\int_0^{C_{n+1}} q_{X_n}(\hat{F}_n(H^-_n)_s)ds}e^{-(t-C_{n+1})}dt\right\}\right]\\
=\ &\mathrm{E}^S_{x_0}\left[I\{H^-_n \in dh\}\zeta_{n+2}(dc \times db|H^-_n, x)\left\{\int_0^{C_{n+1}} \tilde{q}(dx|X_n, \hat{F}_n(H^-_n)_t)e^{-\int_0^t q_{X_n}(\hat{F}_n(H^-_n)_s)ds}dt\right.\right.\\
&\left.\left.+I\{C_{n+1} < \infty\}Q(dx|X_n, B_{n+1})e^{-\int_0^{C_{n+1}} q_{X_n}(\hat{F}_n(H^-_n)_s)ds}\right\}\right]\\
=\ &\hat{\mathrm{E}}^\sigma_{x_0}\left[I\{\hat{H}^-_n \in dh\}\sigma^{(0)}_{n+1}(dc \times db|\hat{H}^-_n, x)\left\{\int_0^{\hat{C}_n} \tilde{q}(dx|X_n, \hat{F}_n(\hat{H}^-_n)_t)e^{-\int_0^t q_{X_n}(\hat{F}_n(\hat{H}^-_n)_s)ds}dt\right.\right.\\
&\left.\left.+I\{\hat{C}_n < \infty\}Q(dx|X_n, \hat{B}_n)e^{-\int_0^{\hat{C}_n} q_{X_n}(\hat{F}_n(\hat{H}^-_n)_s)ds}\right\}\right]\\
=\ &\hat{\mathrm{P}}^\sigma_{x_0}(\hat{H}^-_n \in dh,\ X_{n+1} \in dx,\ \hat{C}_{n+1} \in dc,\ \hat{B}_{n+1} \in db),
\end{aligned}$$

where the third equality is by (20) and (5), the second to the last equality is by the inductive supposition, and the last equality is by (1). It follows that (21) holds for each $n \geq 0$.

To complete the proof of this statement, it remains to take the strategy $\bar{S} = \{\bar{F}_n\}_{n=0}^\infty$ in the standard CTMDP model coming from Proposition 4.1(a), and note that

$$\begin{aligned}
&\mathrm{E}^S_{x_0}\left[\int_0^{\Theta_{n+1}}\int_{\mathbf{A}} c_i^{GO}(X_n, a)F_n(H_n)_t(da)dt\right] = \int_{\mathbf{X}\times\mathbf{A}} c_i^{GO}(x, a)\eta^S_n(dx \times da)\\
=\ &\int_{\mathbf{X}\times\mathbf{A}} c_i^{GO}(x, a)\eta^{\overline{S}}_n(dx \times da) = \mathrm{E}^{\bar{S}}_{x_0}\left[\int_0^{\Theta_{n+1}}\int_{\mathbf{A}} c_i^{GO}(X_n, a)\bar{F}_n(H_n)_t(da)dt\right].
\end{aligned}$$

□

Proof of Theorem 3.3. As mentioned in Remark 3.1, Theorems 3.1 and 3.2 imply that an optimal policy in the gradual-impulsive control problem (3) can be produced from an optimal strategy for the standard CTMDP problem (4) through (10), (13), (14), (16) and (18). By Theorem 6.2 of [19], under the imposed conditions, the standard CTMDP problem (4) has a stationary optimal strategy $\overline{S} = \{\overline{F}\}$, where the meaning of a stationary strategy for the standard CTMDP model is given in Proposition 4.1(b). According to Proposition 4.1(b), we see that the optimal policy produced from $\overline{S} = \{\overline{F}\}$ using (10), (13), (14), (16) and (18) is stationary, as required. □

5 Further extensions

From the proofs of Theorems 3.1, 3.2 and 3.3, we see that the results in Section 3 hold for an arbitrary initial distribution (instead of a fixed initial state) and more general cost rate and functions. For example, instead of J constraints induced by the cost rates and functions $\{c_i^G, c_i^I\}_{i=1}^J$, Theorems 3.1,

3.2 and 3.3 survive if we instead consider an arbitrary family of constraints induced by the family of cost rates and functions $\{c_\alpha^G, c_\alpha^I\}_{\alpha\in\Lambda}$, where Λ is an arbitrary (possibly infinite) index set. Theorems 3.1 and 3.2 also hold when the cost rates and functions further depend on the number of transitions n (either induced by natural jumps or impulsive controls), e.g., one can consider functions $\{l_i^{(n)}\}_{i=0}^J$ in (3) instead of the n-independent functions $\{l_i\}_{i=0}^J$.

Another direction of extension is that we could consider the gradual-impulsive control problem (3) over a class of randomized policies, defined as follows. Let $\mathcal{A}^G = \bigcup_{\hat{a}\in\mathbf{A}^G}\{\rho = \{\rho_t\}_{t\in(0,\infty)} \in \mathcal{R}(\mathbf{A}^G) : \ \rho_t(da) \equiv \delta_{\hat{a}}(da)\}$. We may identify it with $\mathbf{A}^G$. Employing the notations in Definition 2.1, a randomized policy in the gradual-impulsive control model is a sequence $\sigma = (\sigma_n^{(0)}, \sigma_n^{(1)})$, where $\sigma_n^{(1)}(d\rho|\hat{h}_n, \hat{c}, \hat{b}) = \sigma_n^{(1)}(d\rho|x_n)$ is concentrated on $\mathcal{A}^G$. Since $\mathcal{A}^G$ is identified with $\mathbf{A}^G$, we may identify $\sigma_n^{(1)}(d\rho|x_n)$ with a stochastic kernel $\tilde{\varphi}_n(da|x_n)$. A randomized policy is called stationary if $\sigma_n^{(0)}$ is in the form of (7), and $\tilde{\varphi}_n(da|x) = \tilde{\varphi}(da|x)$ for all $n \geq 0$. We say that a randomized policy σ' replicates a policy (in terms of performance measures) if $\hat{W}_i(x_0, \sigma) = \hat{W}_i(x_0, \sigma')$ for all $i = 0, 1, \ldots, J$.

Theorem 5.1 *Consider the gradual-impulsive control problem (3). Under the conditions of Theorem 3.3, there exists an optimal stationary policy, which is the same as the one from Theorem 3.3, and is replicated by a stationary randomized policy, which can be obtained using (10), (13), (14) and (18).*

Proof. The above statement follows from the proof of Theorem 3.1 and the statements of Theorems 3.2 and 3.3. The details are as follows. Firstly, by Theorem 3.1, given any $\overline{S}$ in the standard CTMDP model, there exists a policy $\sigma = \{\sigma_n^{(0)}, \hat{F}_n\}_{n=0}^\infty$, which can be obtained using (10), (13), (14), (16) and (18), and satisfies $\hat{W}_i(x_0, \sigma) = W_i(x_0, \overline{S})$ for all $i = 0, 1, \ldots, J$. By inspecting the proof of Theorem 3.1, for this policy σ, there exists a randomized policy $\sigma' = \{\sigma_n^{(0)}, \tilde{\varphi}_n\}_{n=0}^\infty$, which can be obtained using (10), (13), (14) and (18), and satisfies $\hat{W}_i(x_0, \sigma') = W_i(x_0, \overline{S})$ for all $i = 0, 1, \ldots, J$. Note that when $\overline{S}$ is a stationary strategy[3] in the standard CTMDP model, then the corresponding policy σ and the randomized policy σ' are both stationary[4], too. Secondly, by Theorem 3.2, given any policy σ, by the discussions in the first step, there exist a replicating randomized policy σ'. Finally, one should refer to Theorem 3.3 for the existence of a stationary optimal policy σ. □

Here we mention that a randomized policy is realizable in the sense that it induces a jointly measurable controlling (action) process, whereas a policy does not in general. The interested reader is referred to [30] for detailed discussions on this issue in the context of CTMDP models.

6 Conclusion

In this paper, we showed that a gradual-impulsive control problem for CTMDPs (with the total cost criteria and constraints) can be reduced to a standard CTMDP problem with gradual control only. In doing so, we also obtained a simple class of policies sufficient for solving the gradual-impulsive control problem. The usefulness of this reduction method was then demonstrated by applying it to showing the existence of an optimal stationary policy for the gradual-impulsive control problem under a very natural set of conditions.

Acknowledgement

This work was partially supported by the Royal Society (grant number IE160503).

[3]The term of stationary strategy in the standard CTMDP model was defined in Proposition 4.1(b).

[4]The term of stationary policy was defined in Theorem 3.3, whereas the term of stationary randomized policy was defined in the paragraph above Theorem 5.1.

References

[1] Altman, E. (1999). *Constrained Markov Decision Processes*. Chapman and Hall/CRC, Boca Raton.

[2] Bensoussan, A. and Lions, J.T. (1975). Optimal impulse and continuous control: method of nonlinear quasi-variational inequalities. *Trudy Mat. Inst. Steklov* **134**, 5–22.

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

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

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

[6] de Saporta, B., Dufour, F. and Geeraert, A. (2017). Optimal strategies for impulse control of piecewise deterministic Markov processes. *Automatica* **77**, 219–229.

[7] Dempster, M. and Ye, J. (1995). Impulse control of piecewise deterministic Markov processes. *Ann. Appl. Probab.* **5**, 399–423.

[8] Donnelly, R. and Gan, L. (2018). Optimal decisions in a time prioty queue. *Appl. Math. Finance* **25**, 107–147.

[9] Dufour, F. and Genadot, A. (2019). A convex programming approach for discrete-time Markov decision processes under the expected total reward criterion. Preprint available at arXiv:1903.08853.

[10] Dufour, F. and Piunovskiy, A. (2013). The expected total cost criterion for Markov decision processes under constraints. *Adv. Appl. Probab.* **45**, 837-859.

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

[12] Dufour, F. and Piunovskiy, A. (2016). Impulsive control for continuous-Time Markov decision processes: a linear programming approach. *Appl. Math. Optim.* **74**, 129–161.

[13] Dufour, F., Horiguchi, M. and Piunovskiy, A. (2016). Optimal impulsive control of piecewise deterministic Markov processes. *Stochastics* **88**, 1073–1098.

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

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

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

[17] Gatarek, D. (1992). Optimality conditions for impulse control of piecewise-deterministic processes. *Math. Control Signals Systems* **5**, 217–232.

[18] Guo, X.P. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes: Theory and Applications*. Springer, Heidelberg.

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

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

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

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

[23] Menaldi, J. and Robin, M. (2016). On some optimal stopping problems with constraint. *SIAM J. Control Optim.* **54**, 2650–2671.

[24] Menaldi, J. and Robin, M. (2017). On some impulse control problems with constraint. *SIAM J. Control Optim.* **55**, 3204–3225.

[25] Miller, A., Miller, B. and Stepanyan, K. (2018). Joint continuous and impulsive control of Markov chains. In Proceedings of *26th Mediterranean Conference on Control and Automation*, Zadar, Croatia.

[26] Piunovskiy, A. (2004). Optimal interventions in countable jump Markov processes. *Math. Oper. Res.***29**, 289–308.

[27] Piunovskiy, A. (2004). Multicritieria impulsive control of jump Markov processes. *Math. Meth. Oper. Res.* **60**, 124–144.

[28] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim.* **49**, 2032–2061.

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

[30] Piunovskiy, A. (2018). Realizable strategies in continuous-time Markov decision processes. *SIAM J. Control Optim.* **56**, 473–495.

[31] Plum, H. (1991). Impulsive and continuuosly acting control of jump processes-time discretization. *Stochastics* **36**, 163–192.

[32] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games*. Imperial College Press, London.

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

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

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