Gradual-impulsive control for continuous-time Markov decision processes with total undiscounted costs and constraints

Piunovskiy, Alexey; Zhang, Yi

DOI:
10.1137/21M1444060

SIAM J. CONTROL OPTIM.
Vol. 60, No. 3, pp. 1892–1917

GRADUAL-IMPULSIVE CONTROL FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES WITH TOTAL UNDISCOUNTED COSTS AND CONSTRAINTS: LINEAR PROGRAMMING APPROACH VIA A REDUCTION METHOD*

ALEXEY PIUNOVSKIY† AND YI ZHANG‡

Abstract. We consider the constrained optimal control problem for a continuous-time Markov decision process (CTMDP) with gradual-impulsive control. The performance criteria are the expected total undiscounted costs (from the running cost and the impulsive cost). We justify fully a reduction method, and close an open issue in the previous literature. The reduction method induces an equivalent but simpler standard CTMDP model with gradual control only, based on which, we establish effectively, under rather natural conditions, a linear programming approach for solving the concerned constrained optimal control problem.

Key words. continuous-time Markov decision processes, gradual-impulsive control, linear programming approach, reduction method

AMS subject classifications. 90C40, 60J76

DOI. 10.1137/21M1444060

1. Introduction.

The present paper investigates continuous-time Markov decision processes (CTMDPs) in Borel state and action spaces, where the decision maker can control the process via its local characteristics (transition rate), and also can control directly the state of the process. Such a model is called the gradual-impulsive (control) model. For the gradual-impulsive CTMDP model, we are concerned with the following constrained optimal control problem: the expected total undiscounted cost is to be minimized, subject to other performance measures (objectives) in the same form not exceeding predetermined levels.

The gradual-impulsive CTMDP model is quite general. It has two important submodels. One is the standard CTMDP model, in which the decision maker only controls the transition rate of the process. The other one is the impulsive control model, in which the decision maker can only control instantaneously the state of the process. Each of them has a vast literature: for standard CTMDP models, see the monographs [16, 18, 25] and the more recent one [21], which is influenced by [12, 13]; for the impulsive control model, see e.g., [7, 15, 24]. (The latter references actually dealt with a more general class of processes than what is of concern here, namely, piecewise deterministic processes (PDPs); see [6].) The optimal stopping problem is an important example of impulsive control models, where the decision maker can decide when to stop the process, applying the impulse once and for all; see e.g., [2, 4]. Compared to the aforementioned two submodels, there is relatively less literature on gradual-impulsive CTMDP models; see e.g., [9, 10, 23, 26, 28, 29].

*Received by the editors September 3, 2021; accepted for publication (in revised form) February 28, 2022; published electronically June 27, 2022.
https://doi.org/10.1137/21M1444060
Funding: This paper was discussed at the Liverpool workshop Modern Trends in Controlled Stochastic Processes (July, 2021), supported by the EPSRC (EP/T018216/1).
†Department of Mathematical Sciences, University of Liverpool, Liverpool, Merseyside, L69 7ZL, UK (piunov@liv.ac.uk).
‡Corresponding author. School of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK (y.zhang.29@bham.ac.uk).

Most of the previous literature on gradual-impulsive CTMDP models allows one to apply at most one impulse at a given time moment, and the effect of the impulse is often deterministic, as in the recent work [19]. In the gradual-impulsive CTMDP model considered in this paper an impulse can be applied at any time moment, and one can apply multiple impulses at a single time moment. Such gradual-impulsive CTMDP models were considered in [28, 29] and more recently in [9, 10]. The analysis in [9, 10, 28, 29] as well as in [19] is direct, in the sense that no connection with the standard CTMDP model was explored therein. A different method (the so called time discretization method) was taken in [23, 26], where the gradual-impulsive CTMDP model and the associated optimal control problem were studied as the limit of a sequence of discrete-time problems (for the skeleton models). The skeleton models include complicated transition probabilities.

The present paper differs from the previous literature in terms of the problem statement and the methods of investigation. The concerned optimal control problem for the gradual-impulsive CTMDP considered in the majority of the previous literature is unconstrained (with a single objective), as is the case in [9, 19, 23, 28, 29] as well as in [5, 8, 17]. The main optimality result in these papers was the establishment of the Bellman (optimality) equation, which is used to characterize and show the existence of optimal strategies (often known as the dynamic programming method). One of the relatively few works dealing with constrained problems for gradual-impulsive CTMDP models is [10], where the performance criteria are the expected total discounted costs, and a linear programming approach was established. The linear program formulation in [10] is a consequence of direct investigations of the occupation measures and their characterizations. For their arguments, extra conditions (e.g., bounded transition rate) are needed, and the role of the positive discount factor is important. In this connection, we point out that the discounted problem is a special case of the total undiscounted problem considered in the present paper, and the method of investigation here is quite different from [10], and consequently, we do not need to impose any conditions on the growth of the transition and cost rates. More precisely, our investigation is based on the reduction of the gradual-impulsive CTMDP model to an equivalent but simpler standard CTMDP model. The reduction of the gradual-impulsive model for PDPs to an equivalent model with gradual control only was proposed in [8]. The reduction method in [8] is different from the one proposed here. In greater detail, in PDPs, apart from natural jumps, the process also jumps when it hits the active boundary of the state space. In the gradual control model of PDPs, apart from controlling the transition rate, there is also boundary control, corresponding to when the process hits the active boundary. The target in [8] was to reduce the impulse control to a boundary control in the new model by introducing, along with other extra components, fictitious time in the state of the induced model. The idea is to replicate the original impulse epoch when the fictitious time component in the new model hits a suitably introduced boundary. When applying this method to a CTMDP, the induced model will be a PDP (no longer piecewise constant), with additional boundary control (besides the control of the transition rate) and a much more complicated state space than the original one.

Our main contributions are as follows.

(a) We fully justify that the gradual-impulsive CTMDP model can be reduced to an equivalent and simpler standard CTMDP model with the same state space. This reduction method was partially addressed and justified in [22], where it was assumed that the transition intensities are separated from zero at each state. This condition was essentially used in the argument in [22]. Here we manage to remove this extra condition, which, in our opinion, is a significant improvement. In fact, this turns out to be a delicate issue, and calls for a new and different proof. The proof is based on the investigation of several new classes of control strategies, which can be of independent interest in their own right, and were not considered in [22]. The situation is much simpler if one only deals with strategies in a simple form (e.g., stationary), but we consider general strategies.

(b) We establish the linear programming approach to solving constrained gradual-impulsive optimal control problem for CTMDPs with total undiscounted cost criteria. The linear program formulation itself is interesting, and was not reported in the previous literature, to the best of our knowledge. Moreover, no extra conditions on the growth of the transition and cost rates are needed. This is achieved by referring to the relevant results for the equivalent standard CTMDP problem, and thus also demonstrates the effectiveness of the reduction method fully justified in (a).

The rest of this paper is organized as follows. In section 2 we describe the gradual-impulsive CTMDP model and the standard CTMDP model, and state the constrained optimal control problems under consideration. In section 3 we present the main statements concerning the reduction method as well as the linear programming approach to the constrained optimal control problem. The justification of the reduction method is postponed to section 4, which also introduces some new classes of strategies and the auxiliary statements for them.

2. Model descriptions. In this section, we describe the gradual-impulsive control model $\mathcal{M}$ and the model $\mathcal{M}^{GO}$ with gradual control only, as in [22], which also goes back to [27, 29].

2.1. Gradual-impulsive control model. We describe the primitives of the gradual-impulsive control model $\mathcal{M} =: \{\mathbf{X}, \mathbf{A}^G, \mathbf{A}^I, q, Q, \{c_i^G, c_i^I\}_{i=0}^J\}$ 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. Furthermore, we assume without loss of generality that $\mathbf{A}^G$ and $\mathbf{A}^I$ are two disjoint measurable subsets of a Borel space $\mathbf{A}$ such that $\mathbf{A} = \mathbf{A}^G \cup \mathbf{A}^I$. 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 $(x,a) \in \mathbf{X}\times\mathbf{A}^G$. We introduce the postjump measure $\tilde{q}(dy|x,a) := q(dy\setminus\{x\}|x,a)\ \ \forall\, x \in \mathbf{X},\ a \in \mathbf{A}^G$. 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 $\mathcal{B}(\mathbf{X})$. We assume, without loss of generality that

$$Q(\{x\}|x,b) = 0\ \ \forall\, x \in \mathbf{X},\ b \in \mathbf{A}^I. \tag{2.1}$$

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 (2.3). For each $i \in \{0,1,\dots,J\}$, c_i^G and c_i^I are $[0,\infty]$-valued measurable functions on $\mathbf{X}\times\mathbf{A}^G$ and $\mathbf{X}\times\mathbf{A}^I$, respectively. We emphasize the following.

Remark 2.1. The assumption on the positivity of the cost functions is relevant to optimality results and for brevity. The reduction results (see Theorem 3.1 below), actually hold for $[-\infty,\infty]$-valued c_i^I and c_i^G: one only needs to apply the reduction to the two performance measures induced by the positive and negative parts of c_i^I and c_i^G separately.

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 $\rho^0 = \{\rho_t^0(da)\}_{t\in(0,\infty)} \in \mathcal{R}(\mathbf{A}^G)$. Here, $\mathcal{R}(\mathbf{A}^G)$ 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 $(\mathbf{A}^G, \mathcal{B}(\mathbf{A}^G))$. 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, unless stated otherwise, 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 $\frac{\tilde{q}(dy|x_0,\rho_{t_1}^0)}{q_{x_0}(\rho_{t_1}^0)}$. 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. It is thus natural to embed the gradual-impulsive control problem into a discrete-time Markov decision process (DTMDP) (but with a complicated action space involving the space of relaxed control functions.) This way of describing gradual-impulsive control problems for CTMDPs goes back to Yushkevich [29].

The state space of this DTMDP is $\hat{\mathbf{X}} := \{(\infty, x_\infty)\} \cup [0,\infty) \times \mathbf{X}$, where (∞, x_∞) is an isolated point in $\hat{\mathbf{X}}$. The first coordinate of $\hat{x} = (\theta, x) \in \hat{X}$ represents the previous sojourn time in the gradual-impulsive control model, and the state of the controlled process in the gradual-impulsive control model 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]. An element ρ of $\mathcal{R}(\mathbf{A}^G)$ is understood as a relaxed control function, and its value $\rho_t \in \mathcal{P}(\mathbf{A}^G)$ at $t > 0$ can be understood as the distribution of the gradual control after time duration t from the moment when this relaxed control function is selected. In the case $\rho_t(da)$ does not change with $t > 0$, i.e., the relaxed control function is a constant one, with slight abuse of notation, we often write it as $\rho(da)$, omitting the subscript t. According to Lemma 3 of [27], 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 [27], 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 $\hat{a} = (\hat{c}, \hat{b}, \rho) \in \hat{\mathbf{A}}$,

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

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 [27] and its proof therein.

Similarly, the cost functions $\{l_i\}_{i=0}^J$ defined below are measurable on $\hat{\mathbf{X}}\times\hat{\mathbf{A}}\times\hat{\mathbf{X}}$:

$$l_i((\theta,x),\hat{a},(t,y)) := I\{x\in\mathbf{X}\}\left\{\int_0^t c_i^G(x,\rho_s)ds + I\{t=\hat{c}<\infty\}c_i^I(x,\hat{b})\right\} \tag{2.2}$$

for each $i=0,1,\dots,J$ 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 has been in use, and we have used the following generic notation: for each probability measure μ on $\mathcal{B}(\mathbf{A}^G)$ and measurable function f on $\mathbf{A}^G$, we put $f(\mu):=\int_{\mathbf{A}^G} f(x)\mu(dx)$ 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 $\mathbf{A}^G$. The interpretation is that the pair $(\hat{c},\hat{b})$ is the pair of the planned time duration until the next impulse and the next planned impulse (provided that no natural jump has taken place before then), and ρ is the relaxed control function 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 processes 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 $\{\hat{a}_n\}_{n=0}^\infty$.

DEFINITION 2.1. (a) *A strategy in model $\mathcal{M}$ is given by $\sigma=\{\sigma_n^{(0)},\hat{F}_n\}_{n=0}^\infty$, where for each $n\ge 0$, $\sigma_n^{(0)}(d\hat{c}\times d\hat{b}|\hat{h}_n)$ is a stochastic kernel, and $\hat{F}_n$ is an $\mathcal{R}(\boldsymbol{A}^G)$-valued measurable mapping in $(\hat{h}_n,\hat{c},\hat{b})$, where $\hat{h}_n:=(\hat{x}_0,(\hat{c}_0,\hat{b}_0),\hat{x}_1,(\hat{c}_1,\hat{b}_1),\dots,\hat{x}_n)$.*

(b) *A strategy $\sigma=\{\sigma_n^{(0)},\hat{F}_n\}_{n=0}^\infty$ in $\mathcal{M}$ is called stationary if for each $n\ge 0$,*

$$\sigma_n^{(0)}(d\hat{c}\times d\hat{b}|\hat{h}_n)=\sigma^{S,(0)}(d\hat{c}\times d\hat{b}|x_n),\hat{F}_n(\hat{h}_n,\hat{c},\hat{b})_t(da)=\hat{F}^S(x_n)(da)\quad\forall\ t>0,$$

where $\sigma^{S,(0)}(d\hat{c}\times d\hat{b}|x)$ and $\hat{F}^S(x)(da)$ are stochastic kernels on $\mathcal{B}([0,\infty]\times\boldsymbol{A}^I)$ concentrated on $\{0,\infty\}\times\boldsymbol{A}^I$ and on $\mathcal{B}(\boldsymbol{A}^G)$ given $x\in\boldsymbol{X}$. Here, in line with the interpretation of the element of $\mathcal{R}(\boldsymbol{A}^G)$, $\hat{F}^S(x_n)(da)$ is understood as a constant relaxed control function. We identify such a stationary strategy in $\mathcal{M}$ with $\sigma^S=(\sigma^{S,(0)},\hat{F}^S)$.

Under a strategy σ in $\mathcal{M}$, having in hand $\hat{h}_n$, the decision maker selects $(\hat{c}_n,\hat{b}_n)$ according to $\hat{\sigma}_n^{(0)}(d\hat{c}\times d\hat{b}|\hat{h}_n)$, and after that, chooses the relaxed control function $\rho^n=\hat{F}_n(\hat{h}_n,\hat{c}_n,\hat{b}_n)\in\mathcal{R}(\mathbf{A}^G)$ to be used during the next sojourn time. Note that the class of strategies defined above covers the particular case when one a priori determines a fixed time moment, say T, of applying an impulse: this then corresponds to $\sigma_n^{(0)}(d\hat{c}\times\mathbf{A}^I|\hat{h}_n)=\delta_{T-\hat{t}_n}(d\hat{c})$ provided that $\hat{t}_n\le T$, where $\hat{t}_n=\sum_{i=1}^n\hat{\theta}_n$ is the realized time of the nth jump moment, induced by either natural or active (impulsive) jumps.

Given $\hat{x}_0=(0,x_0)\in\hat{\mathbf{X}}$ and a strategy 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

$$\text{minimize over }\sigma\in\Sigma:\hat{W}_0(x_0,\sigma)\text{ such that }\hat{W}_j(x_0,\sigma)\le d_j,\ j=1,\dots,J, \tag{2.3}$$

where for each $0 \leq j \leq J$, $\hat{W}_j(x_0,\sigma) := \hat{\mathrm{E}}^{\sigma}_{x_0}\left[\sum_{n=0}^{\infty} l_j(\hat{X}_n,\hat{A}_n,\hat{X}_{n+1})\right]$, $\{d_j\}_{j=1}^J \in \mathbb{R}^J$ is a fixed vector of constants, and x_0 is a fixed element of $\mathbf{X}$.

2.2. Standard CTMDP model. In a standard CTMDP model, there is only gradual control. Its system primitives are $\mathcal{M}^{GO} := \{\mathbf{X},\mathbf{A},q^{GO},\{c_i^{GO}\}_{i=0}^J\}$. 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 a collection of $[0,\infty]$-valued 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 function $\rho\in\mathcal{R}(\mathbf{A})$, where $\mathcal{R}(\mathbf{A})$ is understood as $\mathcal{R}(\mathbf{A}^{\mathcal{G}})$ with $\mathbf{A}^G$ being replaced by $\mathbf{A}$, 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 $(x_0,\theta_1,x_1,\dots,\theta_{m-1},x_{m-1},\infty,x_\infty,\infty,x_\infty,\dots)$, 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 $\mathbf{X}_\infty=\mathbf{X}\cup\{x_\infty\}$. 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 $t_\infty:=\lim_{n\to\infty}t_n$. 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 $\overline{S}$ in the standard CTMDP model $\mathcal{M}^{GO}$ is given by $\overline{S}=\{\overline{F}_n\}_{n=0}^\infty$, where for each $n\geq 0$, $\overline{F}_n$ is a measurable mapping on $\boldsymbol{H}_n$ taking values in $\mathcal{R}(\boldsymbol{A})$. It is called Markov if $\overline{F}_n(h_n)=\overline{F}_n^M(x_n)$ for some measurable mapping $\overline{F}_n^M$ from $\boldsymbol{X}$ to $\mathcal{R}(\boldsymbol{A})$. In this case, we identify $\overline{S}$ with $\{\overline{F}_n^M\}_{n\geq 0}=:\overline{S}^M$. A strategy $\overline{S}=\{\overline{F}_n\}_{n\geq 0}$ in $\mathcal{M}^{GO}$ is called stationary if $\overline{F}_n(h_n)_t(da)=\overline{F}^S(x_n)(da)$ for some stochastic kernel $\overline{F}^S(x)(da)$ on $\mathcal{B}(\boldsymbol{A})$ given $x\in\boldsymbol{X}$. In this case, we identify such a stationary strategy $\overline{S}$ with $\overline{F}^S$.*

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 $\Gamma_1\in\mathcal{B}([0,\infty))$, $\Gamma_2\in\mathcal{B}(\mathbf{X})$,

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

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 (2.3):

$$\text{minimize over } \overline{S}: W_0(x_0,\overline{S}) \text{ subject to } W_j(x_0,\overline{S})\leq d_j,\ j=1,\dots,J, \tag{2.4}$$

with $W_i(x_0,\overline{S}) := \mathrm{E}_{x_0}^{\overline{S}}[\sum_{n=0}^{\infty} I\{T_n<\infty\}\int_{T_n}^{T_{n+1}} c_i^{GO}(X_n,\overline{F}_n(H_n)_{t-T_n})dt]$ for each $i = 0,1,\dots,J$ Here, the constants J and $\{d_j\}_{j=1}^J$. are the same as in problem (2.3).

3. Main results.

3.1. Reduction results. In the rest of this paper, we consider the following standard CTMDP model $\mathcal{M}^{GO}$ induced by the gradual-impulsive control model $\mathcal{M}$, defined as follows:

$$\begin{aligned}
&\mathbf{A} := \mathbf{A}^I \cup \mathbf{A}^G;\ q^{GO}(dy|x,a) := q(dy|x,a)\ \ \forall\ (x,a)\in \mathbf{X}\times\mathbf{A}^G;\\
&\tilde{q}^{GO}(dy|x,a) := Q(dy|x,a),\ q_x^{GO}(a) := 1\ \ \forall\ (x,a)\in\mathbf{X}\times\mathbf{A}^I;\\
&c_i^{GO}(x,a) := c_i^G(x,a)\ \ \forall\ (x,a)\in\mathbf{X}\times\mathbf{A}^G;\ c_i^{GO}(x,a) := c_i^I(x,a)\ \ \forall\ (x,a)\in\mathbf{X}\times\mathbf{A}^I.
\end{aligned} \tag{3.1}$$

(Equality (2.1) guarantees that q^{GO} defined in the above is indeed a transition rate.)

One purpose of this paper is to show that the gradual-impulsive control model $\mathcal{M}$ can be reduced to this induced model $\mathcal{M}^{GO}$ in the following sense. We say that the model $\mathcal{M}$ can be reduced to the model $\mathcal{M}^{GO}$ if each strategy in $\mathcal{M}$ is replicated by a strategy in $\mathcal{M}^{GO}$, and each strategy in $\mathcal{M}^{GO}$ is replicated by a strategy in $\mathcal{M}$, where a strategy in a model is said to replicate another strategy in a possibly different model if the performance measures of the two strategies in their respective models coincide. We formulate our first main result as follows.

THEOREM 3.1. *The model $\mathcal{M}$ can be reduced to the model $\mathcal{M}^{GO}$. That is, for each strategy σ in $\mathcal{M}$ (or $\overline{S}$ in $\mathcal{M}^{GO}$), there is some strategy $\overline{S}$ in $\mathcal{M}^{GO}$ (respectively, σ in $\mathcal{M}$) such that $\hat{W}_i(x_0,\sigma) = W_i(x_0,\overline{S}),\ i\in\{0,\dots,J\}$.*

Proof. The proof of this theorem is postponed to section 4. ∎

This reduction issue was partially addressed in [22]. where it was established that any strategy in $\mathcal{M}$ can be replicated by a strategy in $\mathcal{M}^{GO}$; see Theorem 3.2 of [22] therein. The opposite direction is more delicate. The corresponding statement, collected as Proposition 3.2 below, was established in [22] under the following extra condition.

CONDITION 3.1. *For each $x\in\boldsymbol{X}$, there is some $\epsilon>0$ such that $q_x(a)\geq\epsilon>0$ for all $a\in\boldsymbol{A}^G$.*

PROPOSITION 3.2. *Suppose Condition 3.1 is satisfied. Then each strategy in $\mathcal{M}^{GO}$ can be replicated by a strategy in $\mathcal{M}^{GO}$, i.e., for each strategy $\overline{S}$ in $\mathcal{M}^{GO}$, there is a strategy σ in $\mathcal{M}$ such that $\hat{W}_i(x_0,\sigma) = W_i(x_0,\overline{S}),\ i\in\{0,\dots,J\}$.*

Proof. See Theorem 3.1 of [22]. ∎

Let us describe a simple example where Condition 3.1 is not satisfied. Let $\mathbf{X}=\mathbb{Z}$. The state $x\in\{0,1,\dots\}$ represents the number of infected population. The subset $\{0,-1,-2,\dots\}$ can be viewed as a cemetery, but instead of merging them to a single state 0, we keep it in the current form, so that (2.1) is satisfied. Each individual gives an infection rate $\lambda>0$, and one may (gradually) vaccinate the population to reduce this infection rate. Suppose $q(\{x+1\}|x,a) = e^{-a}\lambda x I\{x\geq 0\} = q_x(a)$ for $a\in\mathbf{A}^G := [0,\infty)$. Here $a\in\mathbf{A}^G$ is the intensity of vaccination. One may also remove the virus carriers impulsively (at a certain cost), and so $Q(\{x-b\}|x,b)=1$ with $b\in\mathbf{A}^I :-= \{1,\dots,\overline{x}\}$ for some $\overline{x}\in\{1,2,\dots\}$. One may consider the following cost functions. Let $c_0^G(x,a)$ be a function that increases in x for fixed a, $c_0^I(x,b)$ be an increasing function in b, $c_1^G(x,a)$ be an increasing function in a, and $c_1^I(x,b)=0$.

Note that $q_x(a)=0$ for $x\in\{0,-1,\dots\}$ and $\inf_{a\in\mathbf{A}^G} q_x(a)=0$ for $x\in\{1,2,\dots\}$, so that Condition 3.1 is violated.

The main contribution of this paper lies in showing that Condition 3.1 can be withdrawn from Proposition 3.2, and that removal would also complete the proof of Theorem 3.1. We underline that the argument in the proof of Theorem 3.1 of [22] essentially made use of Condition 3.1. Here we will get over this difficulty by introducing and investigating in section 4 auxiliary and new classes of strategies in $\mathcal{M}$ and in $\mathcal{M}^{GO}$, which can be of independent interest.

The situation is simpler if we only consider stationary strategies in $\mathcal{M}^{GO}$. They can be replicated by stationary strategies in $\mathcal{M}$ as observed in the next statement. Its proof can be done directly without assuming Condition 3.1 or involving auxiliary strategies, though the argument cannot handle the case of general strategies. For this reason we omit the details. (They can be found in the preprint of this paper available at arXiv:2112.02674.)

PROPOSITION 3.3. *Each stationary strategy $\overline{F}^S$ in $\mathcal{M}^{GO}$ is replicated by the stationary strategy $\sigma^S=(\sigma^{S,(0)},\hat{F}^S)$ in $\mathcal{M}$ defined as follows: for each $x\in O$ with $O:=\left\{x\in\boldsymbol{X}:\ \int_{\boldsymbol{A}^G} q_x(a)\overline{F}^S(x)(da)+\overline{F}^S(x)(\boldsymbol{A}^I)>0\right\}$, on $\mathcal{B}(\boldsymbol{A}^I)$,*

$$\sigma^{S,(0)}(\Gamma\times d\hat{b}|x)=0\ \forall\ \Gamma\in\mathcal{B}(0,\infty),$$

$$\sigma^{S,(0)}(\{0\}\times d\hat{b}|x)=\frac{\overline{F}^S(x)(d\hat{b})}{\int_{\boldsymbol{A}^G} q_x(a)\overline{F}^S(x)(da)+\overline{F}^S(x)(\boldsymbol{A}^I)},$$

$$\sigma^{S,(0)}(\{\infty\}\times d\hat{b}|x)=\begin{cases}\frac{\overline{F}^S(x)(d\hat{b})}{\overline{F}^S(x)(\boldsymbol{A}^I)}\frac{\int_{\boldsymbol{A}^G} q_x(a)\overline{F}^S(x)(da)}{\int_{\boldsymbol{A}^G} q_x(a)\overline{F}^S(x)(da)+\overline{F}^S(x)(\boldsymbol{A}^I)} & \text{if } \overline{F}^S(x)(\boldsymbol{A}^I)>0,\\ p^{**}(d\hat{b}) & \text{otherwise,}\end{cases}$$

where p^{**} is an arbitrarily fixed probability measure on $\mathcal{B}(\boldsymbol{A}^I)$;

$$\hat{F}^S(x)(da)=\begin{cases}\frac{\overline{F}^S(x)(da\cap\boldsymbol{A}^G)}{\overline{F}^S(x)(\boldsymbol{A}^G)} & \text{if } \overline{F}^S(x)(\boldsymbol{A}^G)>0,\\ p^{*}(da) & \text{otherwise,}\end{cases}$$

where p^* is an arbitrarily fixed probability measure on $\mathcal{B}(\boldsymbol{A}^G)$; whereas for each $x\in\boldsymbol{X}\setminus O$, $\sigma^{S,(0)}(d\hat{c}\times d\hat{b}|x)=\delta_\infty(d\hat{c})p^{**}(d\hat{b})$, $\hat{F}^S(x)(da)=\overline{F}^S(x)(da)$.

3.2. Optimality results. Theorem 3.1 and Proposition 3.3 assert that if one can obtain a stationary optimal strategy for the standard CTMDP problem (2.4) in the induced model $\mathcal{M}^{GO}$, then a stationary optimal strategy for the gradual-impulsive optimal control problem (2.3) can be constructed. In view of this observation, we present conditions that guarantee the existence of an optimal stationary strategy for problem (2.3), and establish a linear program, solving which, one can produce such an optimal stationary strategy.

CONDITION 3.2. *$\boldsymbol{A}^G$ and $\boldsymbol{A}^I$ are compact; $\{c_i^G\}_{i=0}^J$ 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, and 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.*

Suppose that Condition 3.2 is satisfied in this subsection. Let v^* be the minimal nonnegative lower semicontinuous function on $\mathbf{X}$ satisfying the first equality in

$$v^*(x) = \inf_{a\in\mathbf{A}}\left\{\frac{\sum_{j=0}^J c_j^{GO}(x,a)}{\epsilon+q_x^{GO}(a)} + \frac{\int_{\mathbf{X}} v^*(y)\tilde{q}^{GO}(dy|x,a)+\epsilon v^*(x)}{\epsilon+q_x^{GO}(a)}\right\}$$
$$= \frac{\sum_{j=0}^J c_j^{GO}(x,f^*(x))}{\epsilon+q_x^{GO}(f^*(x))} + \frac{\int_{\mathbf{X}} v^*(y)\tilde{q}^{GO}(dy|x,f^*(x))+\epsilon v^*(x)}{\epsilon+q_x^{GO}(f^*(x))},\ x\in\mathbf{X}$$

(recall $\mathbf{A}=\mathbf{A}^G\cup\mathbf{A}^I$), where f^* is a measurable mapping from $\mathbf{X}$ to $\mathbf{A}$. Note that v^* is actually independent of $\epsilon>0$, and the existence of v^* and f^* is guaranteed under Condition 3.2, according to, e.g., Theorem 4.2.1 of [21] and its proof. Put $\mathbf{R}:=\{x\in\mathbf{X}:\ v^*(x)>0\}$. (The intuitive meaning of $\mathbf{R}^c$ is the part of the state space at which it is optimal to apply f^* in the model $\mathcal{M}^{GO}$; the process will remain there with no cost being incurred. Thus, the nontrivial part is to determine the control in $\mathcal{M}^{GO}$when the process is in $\mathbf{R}$.) Now consider the following linear program:

$$\int_{\mathbf{R}\times\mathbf{A}^G} c_0^G(x,a)\nu(dx\times da)+\int_{\mathbf{R}\times\mathbf{A}^I} c_0^I(x,a)\nu(dx\times da)\to\min_\nu$$
$$\text{s.t. }\int_{\mathbf{A}^G} q_y(a)\nu(dy\times da)+\nu(dx\times\mathbf{A}^I)=\delta_{x_0}(dx)+\int_{\mathbf{R}\times\mathbf{A}^G}\tilde{q}(dx|y,a)\nu(dy\times da)$$
$$+\int_{\mathbf{R}\times\mathbf{A}^I} Q(dx|y,a)\nu(dy\times da), \tag{3.2}$$
$$\int_{\mathbf{R}\times\mathbf{A}^G} c_j^G(x,a)\nu(dx\times da)+\int_{\mathbf{R}\times\mathbf{A}^I} c_j^I(x,a)\nu(dx\times da)\le d_j,\ j\in\{1,2,\dots,J\},$$

ν is a measure on $\mathcal{B}(\mathbf{R}\times\mathbf{A}):\nu(dx\times\mathbf{A})$ is a σ-finite measure on $\mathcal{B}(\mathbf{R})$; $\int_{\mathbf{A}^G} q_x(a)\nu(dx\times da)+\nu(dx\times\mathbf{A}^I)$ is σ-finite on $\mathcal{B}(\mathbf{R})$.

Here the set of σ-finite measures ν on $\mathcal{B}(\mathbf{R}\times\mathbf{A})$ forms a positive cone in a suitable vector space of certain set functions, which can be identified as sequences of finite signed measures on a countable measurable partition of $\mathcal{B}(\mathbf{R}\times\mathbf{A})$. (The partitions may be different for different set functions.)

THEOREM 3.4. *Suppose that Condition* 3.2 *is satisfied, and there is a feasible strategy for problem (*2.3*). Then the following assertions hold.*

(a) *There exists an optimal stationary strategy for problem (*2.3*).*

(b) *Suppose the linear program (*3.2*) has a feasible solution, which is the case if problem (*2.3*) has a feasible strategy with finite value. The linear program (*3.2*) has an optimal solution, say ν^*. Consider the stochastic kernel $\overline{F}^S(x)(da)$ on $\mathcal{B}(\boldsymbol{A})$ given $x\in\boldsymbol{X}$ satisfying $\nu^*(dx\times da)=\nu^*(dx\times\boldsymbol{A})\overline{F}^S(x)(da)$ for each $\Gamma\in\mathcal{B}(\boldsymbol{R})$, and $\overline{F}^S(x)(da)=\delta_{f^*(x)}(da)$ for each $x\in\boldsymbol{X}\setminus\boldsymbol{R}$. (Such a stochastic kernel exists because $\nu^*(dx\times\boldsymbol{A})$ is σ-finite on $\mathcal{B}(\boldsymbol{R})$.) Then the stationary strategy $\sigma^S=(\sigma^{S,(0)},\hat{F}^S)$ defined in terms of $\overline{F}^S$ in Proposition* 3.3 *is optimal for problem (*2.3*).*

Proof. By Theorem 3.1, problem (2.3) can be reduced to the induced standard CTMDP problem (2.4). Statement (a) follows from this reduction, Theorem 4.2.2(b) of [21], and Proposition 3.3. Statement (b) further follows from the proof of Theorem 4.2.2(b) of [21]. ∎

DEFINITION 3.5. *A stationary strategy $\sigma^S=(\sigma^{S,(0)},\hat{F}^S)$ in model $\mathcal{M}$ is called pure stationary if*

$$\sigma^{S,(0)}(d\hat{c}\times d\hat{b}|x)=\delta_{\varphi(x)}(d\hat{c})\delta_{\zeta(x)}(d\hat{b}),\ \hat{F}^S(x)(da)=\delta_{f^S(x)}(da)\quad\forall\ t>0,$$

where φ (or ζ, f^S) is a measurable mapping from $\boldsymbol{X}$ to $\{0,\infty\}$ ($\boldsymbol{A}^I$, $\boldsymbol{A}^G$, respectively). We identify such a pure stationary strategy in $\mathcal{M}$ with (φ,ζ,f^S).

Pure stationary strategies are the easiest for implementation. They are often sufficient for problems with a single objective, i.e., when $J=0$ (see e.g., [9, 17]) or when the model satisfies additional structural conditions, such as being convex or atomless (see e.g., [11, 14]). The next example demonstrates that without further conditions, pure stationary strategies are often not sufficient for the constrained problem (2.3). This is a typical situation in the constrained optimization when $J>0$; see [1].

Example 3.1. Let $\mathbf{X}=\{0,1,2,\dots\}$, $\mathbf{A}^G=\{a\}$, $\mathbf{A}^I=\{b\}$ with $a\neq b$, so that we may put $\mathbf{A}=\{a,b\}=\mathbf{A}^G\cup\mathbf{A}^I$. Let $q_0(a)=1=q(\{1\}|0,a)$, $q_x(a)=0$ for all $x\in\{1,2,\dots\}$, $Q(\{x+1\}|x,b)=1$ for all $x\in\mathbf{X}$. Finally, fix $J=1$, $d_1=1$, $x_0=0$, and consider the cost rates and functions defined by

$$\begin{aligned}
&c_0^G(0,a)=1,\ c_0^G(x,a)=0\ \forall\ x\in\{1,2,\dots\},\\
&c_0^I(x,b)=0,\quad c_1^G(x,a)=0\ \forall\ x\in\{0,1,2,\dots\},\\
&c_1^I(0,b)=2,\ c_1^I(x,b)=0\ \forall\ x\in\{1,2,\dots\}.
\end{aligned}$$

Apparently, since the process is essentially only controlled at the state $x=0$ (once the process leaves the state 0, no further cost will be incurred), as far as the performance of pure stationary strategies is concerned, one only needs to consider pure stationary strategies in the following form: $\sigma^{DS}=(\varphi,\zeta,f^S)$ with $\varphi(0)=0$ and $\sigma^{'DS}=(\varphi',\zeta,f^S)$ with $\varphi'(0)=\infty$. We may compute

$$\hat{W}_0(0,\sigma^{DS})=0,\ \hat{W}_1(0,\sigma^{DS})=2>d_1=1,\quad \hat{W}_0(0,\sigma^{'DS})=1,\ \hat{W}_1(0,\sigma^{'DS})=0.$$

Consequently, σ^{DS} is not feasible for problem (2.3). Now consider $\sigma^S=(\sigma^{S,(0)},\hat{F}^S)$ such that $\sigma^{S,(0)}(\{0\}\times\{b\}|0)=0.5=\sigma^{S,(0)}(\{\infty\}\times\{b\}|0)$. Then one can verify that $\hat{W}_0(0,\sigma^S)=\frac{1}{2}<\hat{W}_0(0,\sigma^{'DS})$, $\hat{W}_1(0,\sigma^S)=1$, which is feasible and strictly outperforms $\sigma^{'DS}$, and thus strictly outperforms any feasible pure stationary strategy.

4. Auxiliary statements and proof of Theorem 3.1.

The proof of Theorem 3.1 takes several steps, and makes use of auxiliary classes of strategies in the model $\mathcal{M}$ and the induced model $\mathcal{M}^{GO}$, which are introduced in separate subsections below.

4.1. Pseudo-Poisson-related strategies in $\boldsymbol{\mathcal{M}^{GO}}$.

In what follows, we fix some strictly positive constant $\lambda>0$. Consider the induced model $\mathcal{M}^{GO}$. Let $\overline{\lambda}(a):=\lambda I\{a\in\mathbf{A}^G\}\ \forall\ a\in\mathbf{A}$. For each $n\geq 0$, consider a sequence of stochastic kernels $\{\overline{p}_{n,k}\}_{k\geq 0}$ on $\mathcal{B}(\mathbf{A})$ given $x\in\mathbf{X}$. By a pseudo-Poisson-related strategy in $\mathcal{M}^{GO}$ we mean $(\overline{\lambda},\{\overline{p}_{n,k}\}_{n,k\geq 0})$. Under such a pseudo-Poisson-related strategy, at the beginning of a sojourn time, when the state is x_n, a marked point process is generated according to $(\overline{\lambda},\{\overline{p}_{n,k}\}_{n,k\geq 0})$, and during the sojourn time, we use actions as the marks and change actions only at the arrival times of that marked point process. We illustrate the implementation of such a pseudo-Poisson-related strategy as follows.

- At the initial time with the initial state x_0, generate $\Phi_0^{(0)}\sim\overline{p}_{0,0}(da|x_0)$ and $\Psi_1^{(0)}\sim Exp(\overline{\lambda}(\Phi_0^{(0)}))$, where $Exp(y)$ represents the exponential distribution with rate $y\in[0,\infty)$; an exponential random variable with rate 0 is ∞. Use $\Phi_0^{(0)}$ as the action during $[0,T_1\wedge\Psi_1^{(0)})$. Recall that T_1 is the first sojourn time in $\mathcal{M}^{GO}$.
- If $\Psi_1^{(0)}<T_1$, then generate $\Phi_1^{(0)}\sim\overline{p}_{0,1}(da|x_0)$, $\Psi_2^{(0)}\sim Exp(\overline{\lambda}(\Phi_1^{(0)}))$. Use $\Phi_1^{(0)}$ as the action during $[\Psi_1^{(0)},T_1\wedge\Psi_2^{(0)})$.

- If $\Psi_1^{(0)} \geq T_1$, then generate $\Phi_0^{(1)} \sim \overline{p}_{1,0}(da|X_1)$ and $\Psi_1^{(1)} \sim Exp(\overline{\lambda}(\Phi_0^{(1)}))$. Recall that X_1 is the state variable in $\mathcal{M}^{GO}$ after the first sojourn time. Use $\Phi_0^{(1)}$ as the action during $[T_1, T_2 \wedge (T_1 + \Psi_1^{(1)}))$.
- And so on.

Below we give a more precise definition of a psuedo-Poisson-related strategy, and introduce notations to be used throughout this subsection. We first fix the canonical sample space of the aforementioned marked point process: $\mathbf{\Xi}^{GO} := [0,\infty) \times \mathbf{A} \times ((0,\infty] \times \mathbf{A})^\infty$ is the countable product. The generic notation for an element of $\mathbf{\Xi}^{GO}$ is $\xi = \{(\psi_n, \alpha_n)\}_{n\geq 0}$. Consider the coordinate random variables (viewing $(\mathbf{\Xi}^{GO}, \mathcal{B}(\mathbf{\Xi}^{GO}))$ as a sample space): for each $\xi = \{(\psi_n, \alpha_n)\}_{n\geq 0} \in \mathbf{\Xi}^{GO}$, $\Psi_n(\xi) := \psi_n$, and $\Phi_n(\xi) := \alpha_n$, and $\tau_n := \sum_{k=0}^n \psi_k$.

DEFINITION 4.1. *A pseudo-Poisson-related policy in $\mathcal{M}^{GO}$ is given by a sequence of stochastic kernels $\overline{S}^P = \{\overline{p}_n(d\xi|x)\}_{n\geq 0}$ on $\mathcal{B}(\mathbf{\Xi}^{GO})$ from $x \in \boldsymbol{X}$, where for each $n \geq 0$ and $x \in \boldsymbol{X}$, under $\overline{p}_n(d\xi|x)$, $\overline{p}_n(\Psi_0 \in dt|x) = \delta_0(dt)$, and the random vectors $(\Phi_0, \Psi_1), (\Phi_1, \Psi_2)), \ldots$ are mutually independent satisfying*

$$\overline{p}_n(\Phi_k \in da|x) =: \overline{p}_{n,k}(da|x) \ \forall\ k \in \{0,1,2,\ldots\},$$
$$\overline{p}_n(\Phi_k \in da,\ \Psi_{k+1} > t|x) = e^{-\overline{\lambda}(a)t}\overline{p}_{n,k}(da|x) \ \ \forall\ k \in \{0,1,2,\ldots\},\ t \in (0,\infty).$$

(Note that Ψ_k may take $+\infty$ with a positive probability under $\overline{p}_n(d\xi|x)$. If $\overline{\lambda}(a) \equiv \lambda$, then $\{\sum_{i=0}^n \Psi_i\}_{n\geq 0}$ forms a standard Poisson point process, justifying the use of the prefix "pseudo" here.)

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

$$\begin{aligned}
&\mathrm{P}_{x_0}^{\overline{S}^P}(\Theta_n \in \Gamma_1,\ X_n \in \Gamma_2|H_{n-1}) = \int_{\mathbf{\Xi}^{GO}} \mathrm{P}_{x_0}^{\overline{S}^P,\xi}(\Theta_n \in \Gamma_1,\ X_n \in \Gamma_2|H_{n-1})\overline{p}_n(d\xi|x)\\
&:= \int_{\mathbf{\Xi}^{GO}} \left\{\int_{\Gamma_1} e^{-\int_0^s q_{X_{n-1}}^{GO,\xi}(t)dt}\tilde{q}^{GO,\xi}(\Gamma_2|X_{n-1},s)ds\right\}\overline{p}_n(d\xi|x),\\
&\mathrm{P}_{x_0}^{\overline{S}^P}(\Theta_n = \infty,\ X_n = x_\infty|H_{n-1}) = \int_{\mathbf{\Xi}^{GO}} \mathrm{P}_{x_0}^{\overline{S}^P,\xi}(\Theta_n = \infty,\ X_n = x_\infty|H_{n-1})\overline{p}_n(d\xi|x)\\
&:= \int_{\mathbf{\Xi}^{GO}} \left\{e^{-\int_0^\infty q_{X_{n-1}}^{GO,\xi}(t)dt}\right\}\overline{p}_n(d\xi|x),
\end{aligned} \tag{4.1}$$

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

$$q^{GO,\xi}(dy|x,s) := \sum_{k=0}^\infty q^{GO}(dy|x,\alpha_k)I\{s \in (\tau_k, \tau_{k+1}]\},$$
$$\tilde{q}^{GO,\xi}(dy|x,s) := \sum_{k=0}^\infty \tilde{q}^{GO}(dy|x,\alpha_k)I\{s \in (\tau_k, \tau_{k+1}]\},$$

and $q_x^{GO,\xi}(s) := \sum_{k=0}^\infty q_x^{GO}(\alpha_k)I\{s \in (\tau_k, \tau_{k+1}]\}$. Let the expectation corresponding to $\mathrm{P}_{x_0}^{\overline{S}^P}$ be denoted as $\mathrm{E}_{x_0}^{\overline{S}^P}$.

The system performance under $\overline{S}^P$ is measured by

$$W_i(x_0,\overline{S}^P) := \mathrm{E}_{x_0}^{\overline{S}^P}\left[\sum_{n=0}^{\infty} I\{X_n\neq x_\infty\}\right.$$
$$\left.\times\int_{\mathbf{\Xi}^{GO}}\int_{(0,\infty]}\int_0^t c_i^{GO,\xi}(X_n,s)ds\mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)\overline{p}_n(d\xi|X_n)\right],$$

where $\mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)$ is defined in (4.1), see the terms inside the parentheses therein, and $c_i^{GO,\xi}(x,s) := \sum_{k=0}^{\infty} c_i^{GO}(x,\alpha_k)I\{s\in(\tau_k,\tau_{k+1}]\}$.

THEOREM 4.2. *Each Markov strategy $\overline{S}^M = \{\overline{F}_n^M\}_{n\geq 0}$ in $\mathcal{M}^{GO}$ can be replicated by a pseudo-Poisson-related strategy $\overline{S}^P$ in $\mathcal{M}^{GO}$.*

Proof. Let some Markov strategy $\overline{S}^M = \{\overline{F}_n^M\}_{n\geq 0}$ in $\mathcal{M}^{GO}$ be given, and define the following $\overline{S}^P = \{\overline{p}_n\}_{n\geq 0}$ by

$$\overline{p}_{n,0}(da|x) := \int_0^\infty e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds}(q_x^{GO}(a)+\overline{\lambda}(a))\overline{F}_n^M(x)_t(da)dt, \tag{4.2}$$

where $(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s) := \int_{\mathbf{A}}(\overline{\lambda}(a)+q_x^{GO}(a))\overline{F}_n^M(x)_s(da)$; and for each $k\geq 1$,

$$\overline{p}_{n,k}(da|x) := \int_0^\infty \frac{\overline{\lambda}(\overline{F}_n^M,w)\left(\int_0^w\overline{\lambda}(\overline{F}_n^M,u)du\right)^{k-1}}{(k-1)!}$$
$$\times\left(\frac{\int_w^\infty e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds}(q_x^{GO}(a)+\overline{\lambda}(a))\overline{F}_n^M(x)_t(da)dt}{\int_0^\infty\frac{\overline{\lambda}(\overline{F}_n^M,w)\left(\int_0^w\overline{\lambda}(\overline{F}_n^M,u)du\right)^{k-1}}{(k-1)!}e^{-\int_0^w(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds}dw}\right)dw \tag{4.3}$$

if the denominator does not vanish, otherwise $\overline{p}_{n,k}(da|x)$ is put to be a fixed probability measure $\overline{p}^*$ on $\mathcal{B}(\mathbf{A})$, concentrated on $\mathbf{A}^I$.

For notational convenience, let us introduce

$$Q_{n,k}(w,x) := \frac{\overline{\lambda}(\overline{F}_n^M,w)\left(\int_0^w\overline{\lambda}(\overline{F}_n^M,u)du\right)^{k-1}}{(k-1)!}e^{-\int_0^w(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds},$$

so that

$$\overline{p}_{n,k}(da|x) := \int_0^\infty \frac{\overline{\lambda}(\overline{F}_n^M,w)\left(\int_0^w\overline{\lambda}(\overline{F}_n^M,u)du\right)^{k-1}}{(k-1)!}$$
$$\times\left(\frac{\int_w^\infty e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds}(q_x^{GO}(a)+\overline{\lambda}(a))\overline{F}_n^M(x)_t(da)dt}{\int_0^\infty Q_{n,k}(w,x)dw}\right)dw.$$

It is useful to observe that if

$$\int_0^\infty Q_{n,k}(w,x)dw$$
$$:= \int_0^\infty\left\{\frac{\overline{\lambda}(\overline{F}_n^M,w)\left(\int_0^w\overline{\lambda}(\overline{F}_n^M,u)du\right)^{k-1}}{(k-1)!}e^{-\int_0^w(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds}\right\}dw \tag{4.4}$$

vanishes for some $k \geq 1$, then so does $\int_0^\infty Q_{n,l}(w,x)dw$ for all $l \in \{1,2,\dots\}$.

First of all, let us verify that

$$\mathrm{P}_{x_0}^{\overline{S}^P}(X_n \in dy) = \mathrm{P}_{x_0}^{\overline{S}^M}(X_n \in dy) \tag{4.5}$$

as follows. The case of $n = 0$ is evident. Suppose it holds for some $n \geq 0$, and let us prove $\mathrm{P}_{x_0}^{\overline{S}^P}(X_{n+1} \in \Gamma | X_n = x) = \mathrm{P}_{x_0}^{\overline{S}^M}(X_{n+1} \in \Gamma | X_n = x)$ for each $x \in \mathbf{X}$ and $\Gamma \in \mathcal{B}(\mathbf{X})$, as follows. Note that

$$\begin{aligned} \mathrm{P}_{x_0}^{\overline{S}^P}(X_{n+1} \in \Gamma | X_n = x) &= \int_{\mathbf{\Xi}^{GO}} \int_0^\infty e^{-\int_0^s q_x^{GO,\xi}(t)dt} \tilde{q}^{GO,\xi}(\Gamma|x,s) ds \overline{p}_n(d\xi|x) \\ &= \sum_{k=0}^\infty \int_{\mathbf{\Xi}^{GO}} \int_{(\tau_k,\tau_{k+1})} \tilde{q}^{GO}(\Gamma|x,\alpha_k) \prod_{i=0}^{k-1} e^{-\psi_{i+1} q_x^{GO}(\alpha_i)} e^{-(s-\tau_k) q_x^{GO}(\alpha_k)} ds \overline{p}_n(d\xi|x) \\ &= \sum_{k=0}^\infty \int_{\mathbf{\Xi}^{GO}} \prod_{i=0}^{k-1} e^{-\psi_{i+1} q_x^{GO}(\alpha_i)} I\{\psi_{i+1} < \infty\} \tilde{q}^{GO}(\Gamma|x,\alpha_k) \int_0^{\psi_{k+1}} e^{-q_x^{GO}(\alpha_k)s} ds \overline{p}_n(d\xi|x). \end{aligned} \tag{4.6}$$

Since $(\Phi_0,\Psi_1),(\Phi_1,\Psi_2),\dots$ are mutually independent under $\overline{p}_n(d\xi|x)$, we see, upon computing the integrals with respect to $\overline{p}_n(d\xi|x)$ in the above, that

$$\begin{aligned} &\mathrm{P}_{x_0}^{\overline{S}^P}(X_{n+1} \in \Gamma | X_n = x) \\ &= \sum_{k=0}^\infty \prod_{i=0}^{k-1} \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,i}(da|x) \int_{\mathbf{A}} \frac{\tilde{q}^{GO}(\Gamma|x,a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,k}(da|x), \end{aligned} \tag{4.7}$$

where we recall that $\overline{\lambda}(a) + q_x^{GO}(a) \geq \min\{1,\lambda\} > 0$ for all $a \in \mathbf{A}$. Let us verify for $k \geq 1$ that

$$\prod_{i=0}^{k-1} \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,i}(da|x) = \int_0^\infty Q_{n,k}(w,x)dw \tag{4.8}$$

as follows. When $k = 1$, the left-hand side can be written as

$$\begin{aligned} &\int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,0}(da|x) \\ &= \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \times \int_0^\infty e^{-\int_0^t (\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds} (q_x^{GO}(a) + \overline{\lambda}(a)) \overline{F}_n^M(x)_t(da)dt \\ &= \int_0^\infty \overline{\lambda}(\overline{F}_n^M,t) e^{-\int_0^t (\overline{\lambda}+q_x^{GO})(\overline{F}_n^M,s)ds} dt = \int_0^\infty Q_{n,1}(w,x)dw, \end{aligned}$$

as desired. (Again, we used here the fact that $\overline{\lambda}(a) + q_x^{GO}(a) \geq \min\{1,\lambda\} > 0$ for all $a \in \mathbf{A}$.)

Now assume that (4.8) holds for some $k \geq 1$, and we now must show that

$$\prod_{i=0}^{k} \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,i}(da|x) = \int_0^\infty Q_{n,k+1}(w,x)dw.$$

The case when the right-hand side vanishes is trivial, because it implies the same for the left-hand side by the definition of $\overline{p}_{n,i}$ (see (4.2) and (4.3)), and the observation below (4.4). Thus, we assume that $\int_0^\infty Q_{n,k+1}(w,x)dw > 0$, which is equivalent to $\int_0^\infty Q_{n,k}(w,x)dw > 0$ for all $k \geq 1$ as was observed below (4.4). Then, by the inductive supposition,

$$\prod_{i=0}^{k} \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,i}(da|x) = \prod_{i=0}^{k-1} \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,i}(da|x)$$

$$\times \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,k}(da|x) = \int_0^\infty Q_{n,k}(w,x)dw \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,k}(da|x).$$

The above expression is equal to

$$\begin{aligned}
&\int_0^\infty Q_{n,k}(w,x)dw \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \int_0^\infty \frac{\overline{\lambda}(\overline{F}_n^M, w)\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k-1}}{(k-1)!} \\
&\times \left(\frac{\int_w^\infty e^{-\int_0^t (\overline{\lambda} + q_x^{GO})(\overline{F}_n^M, s)ds} (q_x^{GO}(a) + \overline{\lambda}(a)) \overline{F}_n^M(x)_t(da)dt}{\int_0^\infty Q_{n,k}(w,x)dw}\right) dw \\
&= \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \int_0^\infty \frac{\overline{\lambda}(\overline{F}_n^M, w)\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k-1}}{(k-1)!} \\
&\times \left(\int_w^\infty e^{-\int_0^t (\overline{\lambda} + q_x^{GO})(\overline{F}_n^M, s)ds} (q_x^{GO}(a) + \overline{\lambda}(a)) \overline{F}_n^M(x)_t(da)dt\right) dw \\
&= \int_0^\infty \left[\int_w^\infty \overline{\lambda}(\overline{F}_n^M, t) e^{-\int_0^t (\overline{\lambda} + q_x^{GO})(\overline{F}_n^M, s)ds} dt\right] \frac{\overline{\lambda}(\overline{F}_n^M, w)\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k-1}}{(k-1)!} dw.
\end{aligned}$$

Integrating by parts the above integral, we may write the previous expression as

$$\begin{aligned}
&\left[\int_w^\infty \overline{\lambda}(\overline{F}_n^M, t) e^{-\int_0^t (\overline{\lambda} + q_x^{GO})(\overline{F}_n^M, s)ds} dt \frac{\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^k}{k!}\right]_0^\infty \\
&+ \int_0^\infty \frac{\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^k}{k!} \overline{\lambda}(\overline{F}_n^M, w) e^{-\int_0^w (\overline{\lambda} + q_x^{GO})(\overline{F}_n^M, s)ds} dw = \int_0^\infty Q_{n,k+1}(w,x)dw,
\end{aligned}$$

where for the equality, one may apply routine analysis based on $\overline{\lambda}(a) + q_x^{GO}(a) \geq \min\{1, \lambda\} > 0$ for all $a \in \mathbf{A}$. This thus proves (4.8) for all $k \geq 1$.

We may substitute (4.8) back into (4.7):

$$\begin{aligned}
&\mathrm{P}_{x_0}^{\overline{S}^P}(X_{n+1} \in \Gamma | X_n = x) = \int_{\mathbf{A}} \frac{\tilde{q}^{GO}(\Gamma|x,a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,0}(da|x) + \sum_{k=1}^\infty \int_0^\infty Q_{n,k}(w,x)dw \\
&\times \int_{\mathbf{A}} \frac{\tilde{q}^{GO}(\Gamma|x,a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,k}(da|x) = \int_0^\infty \tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t) e^{-\int_0^t (\overline{\lambda} + q_x^{GO})(\overline{F}_n^M, s)ds} dt \qquad (4.9) \\
&+ \sum_{k=1}^\infty \int_0^\infty \frac{\overline{\lambda}(\overline{F}_n^M, w)\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k-1}}{(k-1)!} \tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t) \int_w^\infty e^{-\int_0^t (\overline{\lambda} + q_x^{GO})(\overline{F}_n^M, s)ds} dt dw
\end{aligned}$$

with the above equalities being valid whether $\int_0^\infty Q_{n,k}(w,x)dw$ vanishes or not; indeed, if $\int_0^\infty Q_{n,k}(w,x)dw = 0$, then the summands in the last one of the previous equalities vanish, too.

Note that, for $k \geq 1$,

$$\int_0^\infty \frac{\overline{\lambda}(\overline{F}_n^M, w)\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k-1}}{(k-1)!}\tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t)\int_w^\infty e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M, s)ds}dtdw$$
$$= \int_0^\infty \int_0^t \overline{\lambda}(\overline{F}_n^M, w)\left\{\frac{\left(\int_0^w \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k-1}}{(k-1)!}\right\}dw\tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t)e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M, s)ds}dt$$
$$= \int_0^\infty \int_0^t \overline{\lambda}(\overline{F}_n^M, w)\int_{\{0\leq v_1\leq v_2\leq\cdots\leq v_{k-1}\leq w\}}\prod_{j=1}^{k-1}\overline{\lambda}(\overline{F}_n^M, v_j)dv_1dv_2\dots dv_{k-1}dw$$
$$\times\tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t)e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M, s)ds}dt$$
$$= \int_0^\infty \int_{\{0\leq v_1\leq v_2\leq\cdots\leq v_{k-1}\leq w\leq t\}}\prod_{j=1}^{k-1}\overline{\lambda}(\overline{F}_n^M, v_j)\overline{\lambda}(\overline{F}_n^M, w)dv_1dv_2\dots dv_{k-1}dw$$
$$\times\tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t)e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M, s)ds}dt$$
$$= \int_0^\infty \left\{\frac{\left(\int_0^t \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k}}{k!}\right\}\tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t)e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M, s)ds}dt,$$

where the first equality is by the Fubini–Tonelli theorem, and for the second as well as the last equality, recall the following equality, which is valid for any real-valued integrable function f:

$$\left(\int_0^w f(u)du\right)^{k-1} = (k-1)!\int_{\{0\leq v_1\leq v_2\leq\cdots\leq v_{k-1}\leq w\}}\prod_{j=1}^{k-1}f(v_j)dv_1dv_2\dots dv_{k-1}.$$

With the above equalities, (4.9) can be written as follows:

$$\text{(4.10)}\quad \mathrm{P}_{x_0}^{\overline{S}^P}(X_{n+1}\in\Gamma|X_n=x) = \sum_{k=0}^\infty\int_0^\infty\left\{\frac{\left(\int_0^t \overline{\lambda}(\overline{F}_n^M, u)du\right)^{k}}{k!}\right\}\tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t)$$
$$\times e^{-\int_0^t(\overline{\lambda}+q_x^{GO})(\overline{F}_n^M, s)ds}dt = \int_0^\infty \tilde{q}^{GO}(\Gamma|x, \overline{F}_n^M, t)e^{-\int_0^t q_x^{GO}(\overline{F}_n^M, s)ds}dt$$
$$= \mathrm{P}_{x_0}^{\overline{S}^M}(X_{n+1}\in\Gamma|X_n=x),$$

as desired.

The rest verifies

$$\mathrm{E}_{x_0}^{\overline{S}^P}\left[I\{X_n\neq x_\infty\}\int_{\mathbf{\Xi}^{GO}}\int_{(0,\infty]}\int_0^t c_i^{GO,\xi}(X_n,s)ds\mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)\overline{p}_n(d\xi|X_n)\right]$$
$$\text{(4.11)}\qquad = \mathrm{E}_{x_0}^{\overline{S}^M}\left[I\{X_n\neq x_\infty\}\int_0^{\Theta_{n+1}} c_i^{GO}(X_n, \overline{F}_n^M, s)ds\right],$$

which would complete the proof of this theorem. It is sufficient to assume in the rest of this proof that c_i^{GO} is nonnegative and bounded on $\mathbf{X}\times\mathbf{A}$: the general case can be handled based on this simpler case with the help of the monotone convergence theorem.

Note that on $\{X_n\neq x_\infty\}$,

$$\int_{(0,\infty]}\int_0^t c_i^{GO,\xi}(X_n,s)ds\mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)=\mathrm{E}_n^{\overline{S}^P,\xi}\left[\int_0^{\Theta_{n+1}} c_i^{GO,\xi}(X_n,s)ds|X_n\right],$$

where $\mathrm{E}_n^{\overline{S}^P,\xi}[\cdot|X_n]$ is understood with respect to $\mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)$, which is defined in (4.1); see the terms inside the parentheses therein.

Now, the left-hand side of (4.11) can be written as

$$\begin{aligned}
&\mathrm{E}_{x_0}^{\overline{S}^P}\left[I\{X_n\neq x_\infty\}\int_{\mathbf{\Xi}^{GO}}\int_{(0,\infty]}\int_0^t c_i^{GO,\xi}(X_n,s)ds\mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)\overline{p}_n(d\xi|X_n)\right]\\
&=\mathrm{E}_{x_0}^{\overline{S}^P}\left[I\{X_n\neq x_\infty\}\int_{\mathbf{\Xi}^{GO}}\mathrm{E}_{x_0}^{\overline{S}^P,\xi}\left[\int_0^\infty c_i^{GO,\xi}(X_n,s)I\{s<\Theta_{n+1}\}ds|X_n\right]\overline{p}_n(d\xi|X_n)\right]\\
&=\int_{\mathbf{X}}\mathrm{P}_{x_0}^{\overline{S}^P}(X_n\in dx)\left\{\int_{\mathbf{\Xi}^{GO}}\int_0^\infty c_i^{GO,\xi}(x,s)e^{-\int_0^s q_x^{GO,\xi}(t)dt}ds\overline{p}_n(d\xi|x)\right\}.
\end{aligned}$$

Note that the term inside the parenthesis is in the same form as the term on the right-hand side of (4.6), where $\tilde{q}^{GO,\xi}(\Gamma|x,s)$ is replaced by $c_i^{GO,\xi}(x,s)$ with the latter term having been assumed to be nonnegative and bounded. Therefore, the calculations in (4.6)–(4.10) apply with obvious modifications (more precisely, replacing $\tilde{q}^{GO}(\Gamma|x,a)$ by $c_i^{GO}(x,a)$), leading to

$$\begin{aligned}
&\mathrm{E}_{x_0}^{\overline{S}^P}\left[I\{X_n\neq x_\infty\}\int_{\mathbf{\Xi}^{GO}}\int_{(0,\infty]}\int_0^t c_i^{GO,\xi}(X_n,s)ds\mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)\overline{p}_n(d\xi|X_n)\right]\\
&=\int_{\mathbf{X}}\mathrm{P}_{x_0}^{\overline{S}^P}(X_n\in dx)\left\{\sum_{k=0}^\infty\prod_{i=0}^{k-1}\int_{\mathbf{A}}\frac{\overline{\lambda}(a)}{\overline{\lambda}(a)+q_x^{GO}(a)}\overline{p}_{n,i}(da|x)\int_{\mathbf{A}}\frac{c_i^{GO}(x,a)}{\overline{\lambda}(a)+q_x^{GO}(a)}\overline{p}_{n,k}(da|x)\right\}\\
&=\int_{\mathbf{X}}\mathrm{P}_{x_0}^{\overline{S}^P}(X_n\in dx)\left\{\int_0^\infty c_i^{GO}(x,\overline{F}_n^M,t)e^{-\int_0^t q_x^{GO}(\overline{F}_n^M,s)ds}dt\right\}, \qquad (4.12)
\end{aligned}$$

where for the first and the second equality, compare the corresponding terms in the parentheses with (4.7) and (4.10).

On the other hand, the right-hand side of (4.11) can be written as

$$\begin{aligned}
&\mathrm{E}_{x_0}^{\overline{S}^M}\left[I\{X_n\neq x_\infty\}\mathrm{E}_{x_0}^{\overline{S}^M}\left[\int_0^\infty c_i^{GO}(X_n,\overline{F}_n^M,s)I\{s<\Theta_{n+1}\}ds|X_n\right]\right]\\
&=\int_{\mathbf{X}}\mathrm{P}_{x_0}^{\overline{S}^M}(X_n\in dx)\left\{\int_0^\infty c_i^{GO}(x,\overline{F}_n^M,t)e^{-\int_0^t q_x^{GO}(\overline{F}_n^M,s)ds}dt\right\}.
\end{aligned}$$

Since $\mathrm{P}_{x_0}^{\overline{S}^M}(X_n\in dx)=\mathrm{P}_{x_0}^{\overline{S}^P}(X_n\in dx)$ as was verified earlier in this proof (cf. (4.5)), we see that the previous expression coincides with the term on the left-hand side of (4.11), as required. ∎

4.2. Poisson-related strategy in $\mathcal{M}$. Recall that $\lambda \in (0,\infty)$ is a fixed constant. Let $\mathbf{\Xi} := [0,\infty) \times \mathbf{A}^G \times ((0,\infty)\times \mathbf{A}^G)^\infty$ be the countable product. The generic notation for an element of $\mathbf{\Xi}$ is still $\xi = \{(\psi_n,\alpha_n)\}_{n\geq 0}$, and the coordinate random variables are still denoted, for each $\xi = \{(\psi_n,\alpha_n)\}_{n\geq 0} \in \mathbf{\Xi}$, by $\Psi_n(\xi) := \psi_n$ and $\Phi_n(\xi) := \alpha_n$. The context should exclude any confusion with $\mathbf{\Xi}^{GO}$. For each $n \in \{0,1,\dots\}$, let $p_n(d\xi|x)$ be a stochastic kernel on $\mathcal{B}(\mathbf{\Xi})$ given $x \in \mathbf{X}$, which is specified by the following: for each $x \in \mathbf{X}$, under $p_n(d\xi|x)$, the coordinate random variables $\Psi_0, \Phi_0, \Psi_1, \Phi_1, \dots$ are mutually independent and

$$p_n(\Psi_0 \in dt|x) = \delta_0(dt),\ p_n(\Psi_k \leq t|x) = 1 - e^{-\lambda t}\ \ \forall\ k \in \{1,2,\dots\},$$
$$p_n(\Phi_k \in da|x) =: p_{n,k}(da|x)\ \forall\ k \in \{0,1,2,\dots\}.$$

(Hence, we have under $p_n(d\xi|x)$, $\{\sum_{k=0}^n \Psi_k\}_{n\geq 1}$ is a Poisson point process.) Let $\sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi)$ be a stochastic kernel on $\mathcal{B}([0,\infty]\times \mathbf{A}^I)$ from $(x,\xi) \in \mathbf{X}\times\mathbf{\Xi}$.

Definition 4.3. *The pairs $\{(\sigma_n^{P,(0)}, p_n)\}_{n\geq 0} =: \sigma^P$ are called a Poisson-related strategy in $\mathcal{M}$.*

Given $\xi = \{(\psi_n,\alpha_n)\}_{n\geq 0} \in \mathbf{\Xi}$ with the generic notation $\tau_n := \sum_{k=0}^n \psi_k$ for each $n \in \{0,1,\dots\}$, we put

$$q^\xi(dy|x,s) := \sum_{k=0}^\infty q(dy|x,\alpha_k) I\{s \in (\tau_k,\tau_{k+1}]\}, \tag{4.13}$$

$$\tilde{q}^\xi(dy|x,s) := \sum_{k=0}^\infty \tilde{q}(dy|x,\alpha_k) I\{s \in (\tau_k,\tau_{k+1}]\},\quad q_x^\xi(s) := \sum_{k=0}^\infty q_x(\alpha_k) I\{s \in (\tau_k,\tau_{k+1}]\}.$$

Under a Poisson-related strategy $\sigma^P = \{(\sigma_n^{P,(0)}, p_n)\}_{n\geq 0}$, the transition law of $\hat{X}_{n+1} = (\hat{\Theta}_{n+1}, X_{n+1})$ given $\hat{H}_n$ depends on $\hat{H}_n$ only via $(\hat{\Theta}_n, X_n) = (\theta,x) \in \hat{\mathbf{X}}$. It is denoted by $G_n^{\sigma^P}$, and is defined for each bounded measurable function g on $\hat{\mathbf{X}}$ by

$$\begin{aligned}
&\int_{\hat{\mathbf{X}}} g(t,y) G_n^{\sigma^P}(dt\times dy|(\theta,x)) \\
&:= \int_{[0,\infty]\times\mathbf{A}^I\times\mathbf{\Xi}} \left\{ \int_0^{\hat{c}} \int_{\mathbf{X}} g(t,y)\tilde{q}^\xi(dy|x,t) e^{-\int_0^t q_x^\xi(s)ds} dt \right.\\
&\quad + I\{\hat{c}=\infty\} g(\infty,x_\infty) e^{-\int_0^\infty q_x^\xi(s)ds}\\
&\quad \left. + I\{\hat{c}<\infty\} e^{-\int_0^{\hat{c}} q_x^\xi(s)ds} \int_{\mathbf{X}} g(\hat{c},y) Q(dy|x,\hat{b}) \right\} \sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi) p_n(d\xi|x)\\
&=: \int_{[0,\infty]\times\mathbf{A}^I\times\mathbf{\Xi}} \left\{ \int_{\hat{\mathbf{X}}} g(t,y) G_n^{\sigma^P,\xi}(dt\times dy|(\theta,x),\hat{c},\hat{b}) \right\} \sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi) p_n(d\xi|x)
\end{aligned} \tag{4.14}$$

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

Remark 4.1. Note that, $G^{\sigma^P}(dt\times dy|(\theta,x))$ and $G^{\sigma^P,\xi}(dt\times dy|(\theta,x),\hat{c},\hat{b})$, which is defined in (4.14), see the terms inside the parentheses therein, depend on (θ,x) only through $x \in \mathbf{X}_\infty := \mathbf{X}\cup\{x_\infty\}$, and, therefore, we will write $G^{\sigma^P}(dt\times dy|x)$ and $G^{\sigma^P,\xi}(dt\times dy|x,\hat{c},\hat{b})$ for $G^{\sigma^P}(dt\times dy|(\theta,x))$ and $G^{\sigma^P,\xi}(dt\times dy|(\theta,x),\hat{c},\hat{b})$ in what follows. The same applies to $l_i^{\sigma^P,n}(\hat{x}) = l_i^{\sigma^P,n}(x)$ introduced below.

The cost function under σ^P over the corresponding sojourn time is given by

$$l_i^{\sigma^P,n}(x) := \int_{[0,\infty]\times\mathbf{A}^I\times\mathbf{\Xi}} \int_0^\infty I\{x\in\mathbf{X}\}\left\{\int_0^t c_i^{G,\xi}(x,s)ds + I\{t=\hat{c}<\infty\}c_i^I(x,\hat{b})\right\}$$

$$(4.15)\quad \times G_n^{\sigma^P,\xi}(dt\times\mathbf{X}_\infty|x,\hat{c},\hat{b})\sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi)p_n(d\xi|x)\ \forall\ \hat{x}=(\theta,x)\in\hat{\mathbf{X}},$$

where $\mathbf{X}_\infty = \mathbf{X}\cup\{x_\infty\}$, and $c_i^{G,\xi}(x,s) := \sum_{k=0}^\infty c_i^G(x,\alpha_k)I\{s\in(\tau_k,\tau_{k+1}]\}$.

The sequence $\{G_n^{\sigma^P}\}_{n\geq 0}$ together with the initial distribution $\delta_{x_0}(dy)\delta_0(dt)$ defines a probability $\hat{\mathrm{P}}_{x_0}^{\sigma^P}$ on $\left[\bigcup_{n\geq 1}([0,\infty)\times\mathbf{X})^n\times\{(\infty,x_\infty)\}^\infty\right]\cup([0,\infty)\times\mathbf{X})^\infty$. Let $\hat{\mathrm{E}}_{x_0}^{\sigma^P}$ be the expectation with respect to $\hat{\mathrm{P}}_{x_0}^{\sigma^P}$. The system performance under σ^P is measured by

$$\hat{W}_i(x_0,\sigma^P) := \sum_{n\geq 0}\hat{\mathrm{E}}_{x_0}^{\sigma^P}\left[l_i^{\sigma^P,n}(X_n)\right],$$

where we recall the generic notation $\hat{X}_n = (\hat{\Theta}_n, X_n)$ for a state variable in $\mathcal{M}$.

THEOREM 4.4. *Each pseudo-Poisson-related strategy $\overline{S}^P = \{\overline{p}_n\}_{n\geq 0}$ in $\mathcal{M}^{GO}$ can be replicated by a Poisson-related strategy $\sigma^P = \{(\sigma_n^{P,(0)}, p_n)\}_{n\geq 0}$ in $\mathcal{M}$.*

Proof. Let a pseudo-Poisson-related strategy $\overline{S}^P = \{\overline{p}_n\}_{n\geq 0}$ in $\mathcal{M}^{GO}$ be fixed. Consider the Poisson-related strategy $\sigma^P = \{(\sigma_n^{P,(0)}, p_n)\}_{n\geq 0}$ in $\mathcal{M}$ defined by the following: on $\mathcal{B}(\mathbf{A}^G)$, for each $x\in\mathbf{X}$,

$$(4.16)\qquad p_{n,k}(da|x) := \begin{cases} \frac{\overline{p}_{n,k}(da|x)}{\overline{p}_{n,k}(\mathbf{A}^G|x)} & \text{if } \overline{p}_{n,k}(\mathbf{A}^G|x)>0,\\ p^*(da) & \text{otherwise,}\end{cases}$$

where $p^*\in\mathcal{P}(\mathbf{A}^G)$ is fixed; for each $x\in\mathbf{X}$ and $\xi=(\psi_0,\alpha_0,\psi_1,\alpha_1,\dots)\in\mathbf{\Xi}$ with $\tau_n=\sum_{k=0}^n\psi_k$,

$$(4.17)\quad \sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi)$$
$$:= \sum_{k=0}^\infty \delta_{\tau_k}(d\hat{c})\overline{p}_{n,k}(d\hat{b}|x)\prod_{m=0}^{k-1}\overline{p}_{n,m}(\mathbf{A}^G|x) + \delta_\infty(d\hat{c})\prod_{m=0}^\infty \overline{p}_{n,m}(\mathbf{A}^G|x)p^{**}(d\hat{b}),$$

where $p^{**}\in\mathcal{P}(\mathbf{A}^I)$ is a fixed probability measure. Observe that $\sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi)$ defined above depends on $\xi\in\mathbf{\Xi}$ only through $\xi^- := (\psi_0,\psi_1,\psi_2,\dots)$.

In what follows, we will show in two steps that σ^P defined above is a required replicating strategy.

Step 1. First, let us verify that

$$(4.18)\qquad \hat{\mathrm{P}}_{x_0}^{\sigma^P}(X_n\in dy) = \mathrm{P}_{x_0}^{\overline{S}^P}(X_n\in dy).$$

Since the above is clearly valid when $n=0$, both sides being equal to $\delta_{x_0}(dy)$, using an inductive argument, it is sufficient to verify that for an arbitrarily fixed $\Gamma\in\mathcal{B}(\mathbf{X})$ and $x\in\mathbf{X}$, for all $n\geq 0$,

$$(4.19)\qquad G_n^{\sigma^P}([0,\infty)\times\Gamma|x) = \mathrm{P}_{x_0}^{\overline{S}^P}(X_{n+1}\in\Gamma|X_n=x),$$

as follows. (Recall (4.14) and Remark 4.1 for the definition of $G^{\sigma^P,\xi}(dt\times dy|(x,\hat{c},\hat{b})$ with a generic σ^P.)

Recall the right-hand side of (4.19) was computed in (4.7), which can be now written out more explicitly using $\mathbf{A} = \mathbf{A}^I \cup \mathbf{A}^G$, $\mathbf{A}^I \cap \mathbf{A}^G = \emptyset$, $\underline{q}_x^{GO}(a) = q_x(a) I\{a \in \mathbf{A}^G\} + I\{a \in \mathbf{A}^I\}$, $\tilde{q}^{GO}(\Gamma|x,a) = Q(\Gamma|x,a)$ for each $a \in \mathbf{A}^I$, and $\overline{\lambda}(a) = \lambda I\{a \in \mathbf{A}^G\}$ on $\mathbf{A}$:

$$
\begin{aligned}
\mathrm{P}_{x_0}^{\overline{S}^P}(X_{n+1} \in \Gamma | X_n = x) &= \sum_{k=0}^{\infty} \left[\prod_{i=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,i}(da|x) \right] \\
&\times \left(\int_{\mathbf{A}^G} \frac{\tilde{q}(\Gamma|x,a)}{\lambda + q_x(a)} \overline{p}_{n,k}(da|x) + \int_{\mathbf{A}^I} Q(\Gamma|x,a) \overline{p}_{n,k}(da|x) \right) \\
&= \sum_{k=0}^{\infty} \left[\prod_{i=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,i}(da|x) \right] \int_{\mathbf{A}^G} \frac{\tilde{q}(\Gamma|x,a)}{\lambda + q_x(a)} \overline{p}_{n,k}(da|x) \\
&\quad + \sum_{k=0}^{\infty} \left[\prod_{i=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,i}(da|x) \right] \int_{\mathbf{A}^I} Q(\Gamma|x,a) \overline{p}_{n,k}(da|x) \\
&=: B_1 + B_2.
\end{aligned}
\tag{4.20}
$$

On the other hand, the left-hand side of (4.19) may be written as

$$
\begin{aligned}
G_n^{\sigma^P}([0,\infty) \times \Gamma | x) &= \int_{\mathbf{\Xi}} p_n(d\xi|x) \int_{[0,\infty] \times \mathbf{A}^I} \left\{ \int_0^{\hat{c}} \tilde{q}^{\xi}(\Gamma|x,t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \right. \\
&\quad \left. + I\{\hat{c} < \infty\} e^{-\int_0^{\hat{c}} q_x^{\xi}(s)ds} Q(\Gamma|x,\hat{b}) \right\} \sigma_n^{P,(0)}(d\hat{c} \times d\hat{b}|x,\xi) \\
&= \int_{\mathbf{\Xi}} p_n(d\xi|x) \left(\sum_{k=0}^{\infty} \int_0^{\tau_k} \tilde{q}^{\xi}(\Gamma|x,t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \overline{p}_{n,k}(\mathbf{A}^I|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \right. \\
&\quad + \int_0^{\infty} \tilde{q}^{\xi}(\Gamma|x,t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \prod_{m=0}^{\infty} \overline{p}_{n,m}(\mathbf{A}^G|x) \\
&\quad \left. + \sum_{k=0}^{\infty} e^{-\int_0^{\tau_k} q_x^{\xi}(s)ds} \int_{\mathbf{A}^I} Q(\Gamma|x,\hat{b}) \overline{p}_{n,k}(d\hat{b}|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \right),
\end{aligned}
\tag{4.21}
$$

where the first equality is by (4.14), and the second equality is by the above definition of $\sigma_n^{P,(0)}$; see (4.17). Thus,

$$
\begin{aligned}
G_n^{\sigma^P}([0,\infty) \times \Gamma | x) &= \int_{\mathbf{\Xi}} p_n(d\xi|x) \sum_{k=0}^{\infty} \int_0^{\tau_k} \tilde{q}^{\xi}(\Gamma|x,t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \overline{p}_{n,k}(\mathbf{A}^I|x) \\
&\times \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) + \int_{\mathbf{\Xi}} p_n(d\xi|x) \int_0^{\infty} \tilde{q}^{\xi}(\Gamma|x,t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \prod_{m=0}^{\infty} \overline{p}_{n,m}(\mathbf{A}^G|x) \\
&+ \int_{\mathbf{\Xi}} p_n(d\xi|x) \sum_{k=0}^{\infty} e^{-\int_0^{\tau_k} q_x^{\xi}(s)ds} \int_{\mathbf{A}^I} Q(\Gamma|x,\hat{b}) \overline{p}_{n,k}(d\hat{b}|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \\
&=: C_1 + C_2 + C_3.
\end{aligned}
\tag{4.22}
$$

We analyze the above summands term by term as follows.

As for C_3, we see

$$
\begin{aligned}
C_3 &:= \int_{\mathbf{\Xi}} p_n(d\xi|x) \sum_{k=0}^{\infty} e^{-\int_0^{\tau_k} q_x^{\xi}(s)ds} \int_{\mathbf{A}^I} Q(\Gamma|x,\hat{b})\overline{p}_{n,k}(d\hat{b}|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \\
&= \sum_{k=0}^{\infty} \int_{\mathbf{\Xi}} p_n(d\xi|x) \left(\left[\prod_{l=0}^{k-1} e^{-\psi_{l+1} q_x(\alpha_l)} \right] \int_{\mathbf{A}^I} Q(\Gamma|x,\hat{b})\overline{p}_{n,k}(d\hat{b}|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \right) \\
&= \sum_{k=0}^{\infty} \left[\prod_{l=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} p_{n,l}(da|x) \right] \int_{\mathbf{A}^I} Q(\Gamma|x,\hat{b})\overline{p}_{n,k}(d\hat{b}|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \\
&= \sum_{k=0}^{\infty} \left[\prod_{l=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,l}(da|x) \right] \int_{\mathbf{A}^I} Q(\Gamma|x,\hat{b})\overline{p}_{n,k}(d\hat{b}|x) = B_2,
\end{aligned}
$$

where the second to the last equality holds by the definition of $p_{n,l}$: $\overline{p}_{n,l}(da|x) = \overline{p}_{n,l}(\mathbf{A}^G|x)p_{n,l}(da|x)$ (see (4.16)), no matter whether $\prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x)$ vanishes or not, and the same remark applies to the calculations for C_1 and C_2 below, which will not be repeated.

As for C_1, we have

$$
\begin{aligned}
C_1 &= \sum_{k=0}^{\infty} \int_{\mathbf{\Xi}} p_n(d\xi|x) \left(\sum_{l=0}^{k-1} \tilde{q}(\Gamma|x,\alpha_l) \left(\prod_{\nu=0}^{l-1} e^{-\psi_{\nu+1} q_x(\alpha_\nu)} \right) \right. \\
&\quad \left. \times \int_0^{\psi_{l+1}} e^{-t q_x(\alpha_l)} dt \overline{p}_{n,k}(\mathbf{A}^I|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \right) \\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{k-1} \int_{\mathbf{A}^G} \frac{\tilde{q}(\Gamma|x,a)}{q_x(a)+\lambda} p_{n,l}(da|x) \left(\prod_{\nu=0}^{l-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} p_{n,\nu}(da|x) \right) \\
&\quad \times \overline{p}_{n,k}(\mathbf{A}^I|x) \prod_{m=0}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x) \\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{k-1} \int_{\mathbf{A}^G} \frac{\tilde{q}(\Gamma|x,a)}{q_x(a)+\lambda} \overline{p}_{n,l}(da|x) \left(\prod_{\nu=0}^{l-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,\nu}(da|x) \right) \\
&\quad \times (1 - \overline{p}_{n,k}(\mathbf{A}^G|x)) \prod_{m=l+1}^{k-1} \overline{p}_{n,m}(\mathbf{A}^G|x).
\end{aligned}
$$

It is convenient to introduce the following notation:

$$
D_l := \int_{\mathbf{A}^G} \frac{\tilde{q}(\Gamma|x,a)}{q_x(a)+\lambda} \overline{p}_{n,l}(da|x) \left(\prod_{\nu=0}^{l-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,\nu}(da|x) \right).
$$

Then B_1 in (4.20) can be written as

$$
B_1 = \sum_{k=0}^{\infty} \left(\prod_{i=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,i}(da|x) \right) \int_{\mathbf{A}^G} \frac{\tilde{q}(\Gamma|x,a)}{\lambda + q_x(a)} \overline{p}_{n,k}(da|x) = \sum_{l=0}^{\infty} D_l,
$$

which is finite because so is the left-hand side of (4.20).

With the notation of D_l, we now write

$$C_1 = \sum_{k=0}^{\infty}\sum_{l=0}^{k-1} D_l(1-\overline{p}_{n,k}(\mathbf{A}^G|x))\prod_{m=l+1}^{k-1}\overline{p}_{n,m}(\mathbf{A}^G|x).$$

By a similar calculation as for C_1, we may write

$$\begin{aligned}
&C_2 := \int_{\mathbf{\Xi}} p_n(d\xi|x)\int_0^\infty \tilde{q}^{\xi}(\Gamma|x,t)e^{-\int_0^t q_x^{\xi}(s)ds}dt\prod_{m=0}^{\infty}\overline{p}_{n,m}(\mathbf{A}^G|x)\\
&= \sum_{k=0}^{\infty}\int_{\mathbf{A}^G}\frac{\tilde{q}(\Gamma|x,a)}{\lambda+q_x(a)}\overline{p}_{n,k}(da|x)\left(\prod_{\nu=0}^{k-1}\int_{\mathbf{A}^G}\frac{\lambda}{\lambda+q_x(a)}\overline{p}_{n,\nu}(da|x)\right)\prod_{m\geq k+1}\overline{p}_{n,m}(\mathbf{A}^G|x)\\
&= \sum_{k=0}^{\infty} D_k\prod_{m\geq k+1}\overline{p}_{n,m}(\mathbf{A}^G|x).
\end{aligned}$$

Thus, $C_1 + C_2$ equals

$$\begin{aligned}
&\sum_{k=0}^{\infty}\sum_{l=0}^{k-1} D_l(1-\overline{p}_{n,k}(\mathbf{A}^G|x))\prod_{m=l+1}^{k-1}\overline{p}_{n,m}(\mathbf{A}^G|x) + \sum_{k=0}^{\infty} D_k\prod_{m\geq k+1}\overline{p}_{n,m}(\mathbf{A}^G|x)\\
&= \sum_{l=0}^{\infty} D_l\sum_{k=l+1}^{\infty}(1-\overline{p}_{n,k}(\mathbf{A}^G|x))\prod_{m=l+1}^{k-1}\overline{p}_{n,m}(\mathbf{A}^G|x) + \sum_{l=0}^{\infty} D_l\prod_{m\geq l+1}\overline{p}_{n,m}(\mathbf{A}^G|x)\\
&= \sum_{l=0}^{\infty} D_l\left\{\sum_{k\geq l+1}\left(\prod_{m=l+1}^{k-1}\overline{p}_{n,m}(\mathbf{A}^G|x) - \prod_{m=l+1}^{k}\overline{p}_{n,m}(\mathbf{A}^G|x)\right) + \prod_{m\geq l+1}\overline{p}_{n,m}(\mathbf{A}^G|x)\right\}\\
&= \sum_{l=0}^{\infty} D_l = B_1.
\end{aligned}$$

(Recall that $\sum_{l=0}^{\infty} D_l$ converges.) Combining this with the previous observation, we see that $C_1 + C_2 + C_3 = B_1 + B_2$, and by (4.20) and (4.22), we see that (4.19) holds. Consequently, (4.18) follows.

Step 2. In view of the definitions of $\hat{W}_i(x_0,\sigma^P)$ and $W_i(x_0,\overline{S}^P)$, it remains to show that

$$\begin{aligned}
&\hat{\mathrm{E}}_{x_0}^{\sigma^P}[l_i^{\sigma^P,n}(X_n)] = \mathrm{E}_{x_0}^{\overline{S}^P}\Bigg[I\{X_n\neq x_\infty\}\int_{\mathbf{\Xi}^{GO}}\int_{(0,\infty]}\int_0^t c_i^{GO,\xi}(X_n,s)ds\\
&\times \mathrm{P}_n^{\overline{S}^P,\xi}(\Theta_{n+1}\in dt|X_n)\overline{p}_n(d\xi|X_n)\Bigg]
\end{aligned} \tag{4.23}$$

for bounded $[0,\infty)$-valued functions c_i^G, c_i^I, because the general case can be handled using the monotone convergence theorem.

Note that for each $x \in \mathbf{X}$

$$
\begin{aligned}
l_i^{\sigma^P,n}(x) = & \int_{\Xi} \int_{[0,\infty]\times \mathbf{A}^I} \left\{ \int_0^{\hat{c}} \int_0^t c_i^{G,\xi}(x,s) ds q_x^{\xi}(t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \right. \\
& + I\{\hat{c}=\infty\} \int_0^{\infty} c_i^{G,\xi}(x,s) ds e^{-\int_0^{\infty} q_x^{\xi}(s)ds} \\
& \left. + I\{\hat{c}<\infty\} e^{-\int_0^{\hat{c}} q_x^{\xi}(s)ds} \left(\int_0^{\hat{c}} c_i^{G,\xi}(x,s)ds + c_i^I(x,\hat{b}) \right) \right\} \sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi) p_n(d\xi|x) \\
= \lim_{m\to\infty} & \int_{\Xi} \int_{[0,\infty]\times \mathbf{A}^I} \left\{ \int_0^{\hat{c}} \int_0^t c_i^{G,\xi}(x,s) e^{-\frac{s}{m}} ds q_x^{\xi}(t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \right. \\
& + I\{\hat{c}=\infty\} \int_0^{\infty} c_i^{G,\xi}(x,s) e^{-\frac{s}{m}} ds e^{-\int_0^{\infty} q_x^{\xi}(s)ds} + I\{\hat{c}<\infty\} e^{-\int_0^{\hat{c}} q_x^{\xi}(s)ds} \\
& \left. \times \left(\int_0^{\hat{c}} c_i^{G,\xi}(x,s) e^{-\frac{s}{m}} ds + c_i^I(x,\hat{b}) \right) \right\} \sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi) p_n(d\xi|x),
\end{aligned}
$$

where the first equality is by (4.15) and (4.14). Applying, legitimately, integration by parts, we see

$$
\begin{aligned}
& \int_0^{\hat{c}} \int_0^t c_i^{G,\xi}(x,s) e^{-\frac{s}{m}} ds q_x^{\xi}(t) e^{-\int_0^t q_x^{\xi}(s)ds} dt \\
= & \int_0^{\hat{c}} c_i^{G,\xi}(x,t) e^{-\frac{t}{m}} e^{-\int_0^t q_x^{\xi}(s)ds} dt - e^{-\int_0^{\hat{c}} q_x^{\xi}(s)ds} \int_0^{\hat{c}} e^{-\frac{s}{m}} c_i^{G,\xi}(x,s) ds,
\end{aligned}
$$

where all the terms are finite, $\hat{c}$ being finite or not, because so are c_i^I, c_i^G assumed. Substituting the previous equality back into the above formula and applying the monotone convergence theorem, we see

$$
\begin{aligned}
l_i^{\sigma^P,n}(x) = & \int_{\Xi} \int_{[0,\infty]\times \mathbf{A}^I} \left\{ \int_0^{\hat{c}} c_i^{G,\xi}(x,t) e^{-\int_0^t q_x^{\xi}(s)ds} dt + I\{\hat{c}<\infty\} e^{-\int_0^{\hat{c}} q_x^{\xi}(s)ds} c_i^I(x,\hat{b}) \right\} \\
& \times \sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi) p_n(d\xi|x). \qquad (4.24)
\end{aligned}
$$

Observe that the term inside the parenthesis in the above expression is in the same form as the one in the first equality of (4.21), where $\tilde{q}^{\xi}(\Gamma|x,t)$ and $Q(\Gamma|x,\hat{b})$ are now replaced with $c_i^{G,\xi}(x,t)$ and $c_i^I(x,\hat{b})$, respectively. Therefore, by repeating the calculations below (4.21) in Step 1 with obvious modifications, we see that the following equality holds, which corresponds to (4.19) (or more precisely, the established equality $C_1 + C_2 + C_3 = B_1 + B_2$; see more explanations below):

$$
\begin{aligned}
l_i^{\sigma^P,n}(x) = & \sum_{k=0}^{\infty} \left(\prod_{i=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,i}(da|x) \right) \\
& \times \left(\int_{\mathbf{A}^G} \frac{c_i^G(x,a)}{\lambda + q_x(a)} \overline{p}_{n,k}(da|x) + \int_{\mathbf{A}^I} c_i^I(x,a) \overline{p}_{n,k}(da|x) \right).
\end{aligned}
$$

Indeed, the term on the right-hand side of the above equality corresponds to the term on the right-hand side of the first equality in (4.20), which coincides with the right-hand side of (4.19), whereas it was observed earlier that $l_i^{\sigma^P,n}(x)$ corresponds to the left-hand side of (4.19).

Consequently, the left-hand side of (4.23) reads

$$\begin{aligned}&\hat{\mathrm{E}}_{x_0}^{\sigma^P}[l_i^{\sigma^P,n}(X_n)] = \int_{\mathbf{X}} \hat{\mathrm{P}}_{x_0}^{\sigma^P}(X_n \in dx) l_i^{\sigma^P,n}(x) = \int_{\mathbf{X}} \hat{\mathrm{P}}_{x_0}^{\sigma^P}(X_n \in dx)\\ &\times \left\{ \sum_{k=0}^{\infty} \left(\prod_{i=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,i}(da|x) \right) \right.\\ &\left.\times \left(\int_{\mathbf{A}^G} \frac{c_i^G(x,a)}{\lambda + q_x(a)} \overline{p}_{n,k}(da|x) + \int_{\mathbf{A}^I} c_i^I(x,a) \overline{p}_{n,k}(da|x) \right) \right\}.\end{aligned}$$

On the other hand, by (4.12), we may write the right-hand side of (4.23) as

$$\begin{aligned}&\int_{\mathbf{X}} \mathrm{P}_{x_0}^{\overline{S}^P}(X_n \in dx) \left\{ \sum_{k=0}^{\infty} \left(\prod_{i=0}^{k-1} \int_{\mathbf{A}} \frac{\overline{\lambda}(a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,i}(da|x) \right) \right.\\ &\left.\times \int_{\mathbf{A}} \frac{c_i^{GO}(x,a)}{\overline{\lambda}(a) + q_x^{GO}(a)} \overline{p}_{n,k}(da|x) \right\}\\ &= \int_{\mathbf{X}} \mathrm{P}_{x_0}^{\overline{S}^P}(X_n \in dx) \left\{ \sum_{k=0}^{\infty} \left(\prod_{i=0}^{k-1} \int_{\mathbf{A}^G} \frac{\lambda}{\lambda + q_x(a)} \overline{p}_{n,i}(da|x) \right) \right.\\ &\left.\times \left(\int_{\mathbf{A}^G} \frac{c_i^G(x,a)}{\lambda + q_x(a)} \overline{p}_{n,k}(da|x) + \int_{\mathbf{A}^I} c_i^I(x,a) \overline{p}_{n,k}(da|x) \right) \right\},\end{aligned}$$

where the equality is by the definitions of $\overline{\lambda}(a)$, $\mathbf{A}$, c_i^{GO}, and q^{GO}. In view of (4.18), which was established in the above, we see from the previous equality that (4.23) holds, as desired. ∎

4.3. Proof of Theorem 3.1.

Proof. In view of the discussions below Proposition 3.2, we only need to show that each strategy $\overline{S}$ in $\mathcal{M}^{GO}$ can be replicated by a strategy in $\mathcal{M}$.

According to Theorem 2 of [20] (or Theorem 4.1.1 of [21]), for each strategy $\overline{S}$ in $\mathcal{M}^{GO}$, there is a replicating Markov strategy $\overline{S}^M$ in the same model $\mathcal{M}^{GO}$ (recall Definition 2.2). Theorems 4.2 and 4.4 imply that the Markov strategy $\overline{S}^M$ in $\mathcal{M}^{GO}$ is replicated by a Poisson-related strategy σ^P in $\mathcal{M}$. To complete the proof of the statement, it remains to show that this replicating Poisson-related strategy σ^P in $\mathcal{M}$ can be replicated by an (ordinary) strategy σ in the same model $\mathcal{M}$. This is justified as follows. Without loss of generality, we assume that c_i^G and c_i^I are nonnegative and bounded in this proof.

Let some Poisson-related strategy $\sigma^P = \{(\sigma_n^P, p_n)\}_{n\geq 0}$ in $\mathcal{M}$ be fixed. Let $\sigma_n^{(0)}(d\hat{c} \times d\hat{b}|x) := \int_{\mathbf{\Xi}} \sigma_n^{P,(0)}(d\hat{c} \times d\hat{b}|x,\xi) p_n(d\xi|x)$. Then, by Proposition 7.27 of [3] (or Proposition B.1.33 of [21]), there is a stochastic kernel $\hat{p}_n(d\xi|x,\hat{c},\hat{b})$ on $\mathcal{B}(\mathbf{\Xi})$ given $(x,\hat{c},\hat{b}) \in \mathbf{X} \times [0,\infty] \times \mathbf{A}^I$ satisfying

$$\sigma_n^{P,(0)}(d\hat{c} \times d\hat{b}|x,\xi) p_n(d\xi|x) = \hat{p}_n(d\xi|x,\hat{c},\hat{b}) \sigma_n^{(0)}(d\hat{c} \times d\hat{b}|x). \tag{4.25}$$

We define a strategy $\sigma = (\sigma_n^{(0)}, \hat{F}_n)_{n\geq 0}$ in $\mathcal{M}$ as follows. Let $\sigma_n^{(0)}(d\hat{c} \times d\hat{b}|\hat{h}_n) := \sigma_n^{(0)}(d\hat{c} \times d\hat{b}|x_n)$ (Recall the generic notation $\hat{x}_n = (\hat{\theta}_n, x_n)$ for the state in the model

$\mathcal{M}$.) Let

$$\hat{F}_n(\hat{h}_n,\hat{c},\hat{b})_t(da) := \frac{\int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x_n}(u)du} \sum_{k\geq 0}\delta_{\alpha_k}(da) I\{\tau_k < t \leq \tau_{k+1}\}\hat{p}_n(d\xi|x_n,\hat{c},\hat{b})}{\int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x_n}(u)du}\hat{p}_n(d\xi|x_n,\hat{c},\hat{b})}$$
$$=: \hat{F}_n(x_n,\hat{c},\hat{b}),$$

where the generic notations $\xi = \{(\psi_n,\alpha_n)\}_{n\geq 0}\in\mathbf{\Xi}$ and $\tau_k = \sum_{i=0}^k \psi_i$ are in use.

We will show that

$$\hat{\mathrm{P}}^{\sigma}_{x_0}(X_n\in dx) = \hat{\mathrm{P}}^{\sigma^P}_{x_0}(X_n\in dx)\ \forall\ n\geq 0. \tag{4.26}$$

(Recall the generic notation $\hat{X}_n = (\hat{\Theta}_n, X_n)$ in the model $\mathcal{M}$.) Since the initial states are the same, with an inductive argument, it is sufficient to show for $\Gamma\in\mathcal{B}(\mathbf{X})$ and $x\in\mathbf{X}$,

$$\hat{\mathrm{P}}^{\sigma}_{x_0}(X_{n+1}\in\Gamma|X_n=x) = \hat{\mathrm{P}}^{\sigma^P}_{x_0}(X_{n+1}\in\Gamma|X_n=x)\ \forall\ n\geq 1. \tag{4.27}$$

Then,

$$\begin{aligned}
&\tilde{q}(\Gamma|x,\hat{F}_n(x,\hat{c},\hat{b})_t)\\
&= \int_{\mathbf{A}^G}\tilde{q}(\Gamma|x,a)\frac{\int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x}(u)du} \sum_{k\geq 0}\delta_{\alpha_k}(da) I\{\tau_k < t \leq \tau_{k+1}\}\hat{p}_n(d\xi|x,\hat{c},\hat{b})}{\int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})}\\
&= \frac{\int_{\mathbf{\Xi}}\tilde{q}^{\xi}(\Gamma|x,t)e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})}{\int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})};
\end{aligned}$$

recall (4.13) for the definition of $\tilde{q}^{\xi}$. Applying the above equality to $\Gamma=\mathbf{X}$, we see

$$\begin{aligned}
q_x(\hat{F}_n(x,\hat{c},\hat{b})_t) &= \frac{\int_{\mathbf{\Xi}} q^{\xi}_x(t)e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})}{\int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})}\\
&= -\frac{d}{dt}\ln\int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})
\end{aligned}$$

for almost all t, and thus $e^{-\int_0^t q_x(\hat{F}_n(x,\hat{c},\hat{b})_s)ds} = \int_{\mathbf{\Xi}} e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})$. Now,

$$\begin{aligned}
&\hat{\mathrm{P}}^{\sigma}_{x_0}(X_{n+1}\in\Gamma|X_n=x) = \int_{[0,\infty]\times\mathbf{A}^I}\left\{\int_0^{\hat{c}}\tilde{q}(\Gamma|x,\hat{F}_n(x,\hat{c},\hat{b}))_t)e^{-\int_0^t q_x(\hat{F}_n(x,\hat{c},\hat{b})_s)ds}dt\right.\\
&\quad\left. + I\{\hat{c}<\infty\}e^{-\int_0^{\hat{c}} q_x(\hat{F}_n(x,\hat{c},\hat{b})_s)ds}Q(\Gamma|x,\hat{b})\right\}\sigma^{(0)}_n(d\hat{c}\times d\hat{b}|x)\\
&= \int_{[0,\infty]\times\mathbf{A}^I}\left\{\int_0^{\hat{c}}\int_{\mathbf{\Xi}}\tilde{q}^{\xi}(\Gamma|x,t)e^{-\int_0^t q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})dt\right.\\
&\quad\left.+I\{\hat{c}<\infty\}\int_{\mathbf{\Xi}}e^{-\int_0^{\hat{c}} q^{\xi}_{x}(u)du}\hat{p}_n(d\xi|x,\hat{c},\hat{b})Q(\Gamma|x,\hat{b})\right\}\sigma^{(0)}_n(d\hat{c}\times d\hat{b}|x)\\
&= \int_{\mathbf{\Xi}}\int_{[0,\infty]\times\mathbf{A}^I}\left\{\int_0^{\hat{c}}\tilde{q}^{\xi}(\Gamma|x,t)e^{-\int_0^t q^{\xi}_{x}(u)du}dt + I\{\hat{c}<\infty\}e^{-\int_0^{\hat{c}} q^{\xi}_{x}(u)du}Q(\Gamma|x,\hat{b}))\right\}\\
&\quad\times\sigma^{P,(0)}_n(d\hat{c}\times d\hat{b}|x,\xi)p_n(d\xi|x) = \hat{\mathrm{P}}^{\sigma^P}_{x_0}(X_{n+1}\in\Gamma|X_n=x),
\end{aligned} \tag{4.28}$$

where the second to the last equality is by (4.25), and for the last equality, see (4.14). Thus, (4.27) is verified, and (4.26) follows.

Finally, one can show with a similar argument as for (4.24) that

$$\hat{\mathrm{E}}^{\sigma}_{x_0}\left[l_i(\hat{X}_n,\hat{A}_n,\hat{X}_{n+1})|X_n=x\right]=\int_{[0,\infty]\times\mathbf{A}^I}\left\{\int_0^{\hat{c}} c_i^G(x,\hat{F}_n(x,\hat{c},\hat{b})_t)\right.$$

$$\left.\times\, e^{-\int_0^t q_x(\hat{F}_n(x,\hat{c},\hat{b})_s)ds}dt+I\{\hat{c}<\infty\}e^{-\int_0^{\hat{c}} q_x(\hat{F}_n(x,\hat{c},\hat{b})_s)ds}c_i^I(x,\hat{b})\right\}\sigma_n^{(0)}(d\hat{c}\times d\hat{b}|x),$$

where l_i was defined by (2.2). Having inspected that the term in the parenthesis of the last equality is in the same form as the term on the right-hand side of the first equality in (4.28), we see now

$$\hat{\mathrm{E}}^{\sigma}_{x_0}\left[l_i(\hat{X}_n,\hat{A}_n,\hat{X}_{n+1})|X_n=x\right]=\int_{\mathbf{\Xi}}\int_{[0,\infty]\times\mathbf{A}^I}\left\{\int_0^{\hat{c}} c_i^{G,\xi}(x,t)e^{-\int_0^t q_x^{\xi}(u)du}dt\right.$$

$$\left.+\,I\{\hat{c}<\infty\}e^{-\int_0^{\hat{c}} q_x^{\xi}(u)du}c_i^I(x,\hat{b}))\right\}\sigma_n^{P,(0)}(d\hat{c}\times d\hat{b}|x,\xi)p_n(d\xi|x)=l_i^{\sigma^P,n}(x),$$

where the first equality corresponds to the second to the last equality in (4.28), and the last equality holds by (4.24). The previous equality and (4.26) imply that

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

for all $n\geq 0$. The statement is thus proved. ◻

Acknowledgments. This paper was discussed at the Liverpool workshop Modern Trends in Controlled Stochastic Processes (July, 2021). We also thank the associate editor and referees for helpful remarks.

REFERENCES

[1] E. Altman (1999), *Constrained Markov Decision Processes*, Chapman and Hall/CRC Press, Boca Raton, FL.
[2] N. Bäuerle and A. Popp (2018), *Risk-sensitive stopping problems for continuous-time Markov chains*, Stochastics, 90, pp. 411–431.
[3] D. Bertsekas and S. Shreve (1978), *Stochastic Optimal Control*, Academic Press, New York.
[4] O. Costa and M. Davis (1988), *Approximations for optimal stopping of a piecewise-deterministic process*, Math. Control Signals Systems, 1, pp. 123–146.
[5] O. Costa and C. Raymundo (2000), *Impulse and continuous control of piecewise deterministic Markov processes*, Stochastics, 70, pp. 75–107.
[6] M. Davis (1993), *Markov Models and Optimization*, Chapman and Hall, London.
[7] B. de Saporta, F. Dufour, and A. Geeraert (2017), *Optimal strategies for impulse control of piecewise deterministic Markov processes*, Automatica J. IFAC, 77, pp. 219–229.
[8] M. Dempster, and J. Ye (1995), *Impulse control of piecewise deterministic Markov processes*, Ann. Appl. Probab., 5, pp. 399–423.
[9] F. Dufour and A. Piunovskiy (2015), *Impulsive control for continuous-time Markov decision processes*, Adv. Appl. Probab., 47, pp. 106–127.
[10] F. Dufour and A. Piunovskiy (2016), *Impulsive control for continuous-time Markov decision processes: A linear programming approach*, Appl. Math. Optim., 74, pp. 129–161.
[11] F. Dufour and R. Stockbridge (2012), *On the existence of strict optimal controls for constrained, controlled Markov processes in continuous time*, Stochastics, 84, pp. 55–78.
[12] E. Feinberg (2004), *Continuous time discounted jump Markov decision processes: A discrete-event approach*, Math. Oper. Res., 29, pp. 492–524.

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

[14] E.A. Feinberg and A. Piunovskiy (2019), *Sufficiency of deterministic policies for atomless discounted and uniformly absorbing MDPs with multiple criteria*, SIAM J. Control Optim., 57, pp. 163–191.

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

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

[17] X. Guo, A. Kurushima, A. Piunovskiy, and Y. Zhang (2021), *On gradual-impulse control of continuous-time Markov decision processes with exponential utility*, Adv. Appl. Probab., 53, pp. 301–334.

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

[19] A. Miller, B. Miller, and K. Stepanyan (2020), *Simultaneous impulse and continuous control of a Markov chain in continuous time*, Autom. Remote Control, 81, pp. 469–482.

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

[21] A. Piunovskiy and Y. Zhang (2020), *Continuous-Time Markov Decision Processes*, Springer, Cham, Switzerland.

[22] A. Piunovskiy and Y. Zhang (2020), *On reducing a constrained gradual-impulsive control problem for a jump Markov model to a model with gradual control only*, SIAM J. Control. Optim., 58, pp. 192–214.

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

[24] E. Presman and S. Sethi (2006), *Inventory models with continuous and Poisson demands and discounted and average costs*, Prod. Oper. Manag., 15, pp. 279–293.

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

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

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

[28] A. Yushkevich (1983), *Continuous time Markov decision processes with interventions*, Stochastics, 9, pp. 235–274.

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