Adv. Appl. Prob. **47,** 106–127 (2015)
Printed in Northern Ireland

IMPULSIVE CONTROL FOR CONTINUOUS-TIME MARKOV DECISION PROCESSES

FRANÇOIS DUFOUR,* *Université Bordeaux, IMB and INRIA Bordeaux Sud-Ouest*
ALEXEI B. PIUNOVSKIY,** *University of Liverpool*

Abstract

In this paper our objective is to study continuous-time Markov decision processes on a general Borel state space with both impulsive and continuous controls for the infinite time horizon discounted cost. The continuous-time controlled process is shown to be nonexplosive under appropriate hypotheses. The so-called Bellman equation associated to this control problem is studied. Sufficient conditions ensuring the existence and the uniqueness of a bounded measurable solution to this optimality equation are provided. Moreover, it is shown that the value function of the optimization problem under consideration satisfies this optimality equation. Sufficient conditions are also presented to ensure on the one hand the existence of an optimal control strategy, and on the other hand the existence of a ε-optimal control strategy. The decomposition of the state space into two disjoint subsets is exhibited where, roughly speaking, one should apply a gradual action or an impulsive action correspondingly to obtain an optimal or ε-optimal strategy. An interesting consequence of our previous results is as follows: the set of strategies that allow interventions at time $t = 0$ and only immediately after natural jumps is a sufficient set for the control problem under consideration.

Keywords: Impulsive control; continuous control; continuous-time Markov decision process; discounted cost

2010 Mathematics Subject Classification: Primary 90C40
Secondary 60J25

1. Introduction

Continuous-time Markov decision processes (CTMDPs) form a general class of controlled stochastic processes. These are suitable for formulating many optimization problems arising in such applications as engineering, computer science, telecommunications, finance, etc. The analysis of CTMDPs started in the late 1950s and the early 1960s with the pioneering works of Bellman, Blackwell, Howard, and Veinott, to name just a few authors; see, e.g. [1] and [14]. The analysis has mostly concentrated on control problems where the actions influence the transition rate of the process continuously in time. This is nowadays a very active area of research from the point of view of its theoretical foundations, as well as from the applications perspective; see, e.g. the recent books [8] and [18], and the survey [9].

Another class of models with impulsive actions, when the state of the process can be changed instantly, received very little attention. The first attempt to study such problems was probably

Received 1 October 2013; revision received 17 February 2014.
* Postal address: INRIA Bordeaux Sud-Ouest, 200 Avenue de la Vieille Tour, 33405 Talence cedex, France.
Email address: dufour@math.u-bordeaux1.fr
** Postal address: Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK.
Email address: piunov@liverpool.ac.uk

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

due to De Leve [6], [7]. In the 1980s, a systematic study of impulsive control of continuous-time MDPs, including a deterministic drift between the jumps, were conducted on the one hand by Hordijk and van der Duyn Schouten, and on the other hand by Yushkevich. Hordijk and van der Duyn Schouten [11], [12], [13], [19] considered the case where only one impulsive action at a time is permitted. Given an observed history, the planned time moment for the next impulse was deterministic. In these papers, the optimization was performed within a special class of so-called regular and conservative policies. One drawback of this approach is that the use of the dynamic programming method becomes problematic. Yushkevich [20], [21], [22], [23] introduced a new class of stochastic models, the so-called T-processes, which are, roughly speaking, indexed by a parameter representing the natural current time and the number of impulsive actions at that time moment. The introduction of this new family of processes was motivated mainly by the fact that it allows us to consider models with multiple impulses at the same time moment. For a general control model, Yushkevich typically studied the value functions of such control problems in terms of the related quasivariational inequalities. We should also mention another class of controlled models closely related to CTMDPs and called piecewise-deterministic Markov processes for which impulsive control has also been considered. Without attempting to present an exhaustive panorama, we refer the interested reader to the book [5], and the references therein to get a rather complete view of this class of processes.

It is important to point out that impulsive control models are not mentioned in the recent monographs and surveys on CTMDPs [8], [9], [18]. However, they appear naturally in many real-life situations.

The main difficulty in dealing with the impulsive control model is that impulsive actions give rise to a nonstandard path for the controlled process. Indeed, the process may take several different values at the same time moment. This important property makes the classical theory of CTMDPs inapplicable.

In the paper our objective is to develop a new approach to CTMDPs on a general Borel state space $\boldsymbol{X}$ with both impulsive and continuous controls. In our framework, the continuous control influences the intensity of jumps q at all times. This is in opposition to the impulsive control that intervenes by moving the process to a new point of the state space $\boldsymbol{X}$ at some moment specified by the controller. In this context, continuous actions, also called gradual actions by Yushkevich (see, e.g. [21]), take values in the space $\boldsymbol{A}^{\mathrm{g}}$ and lead to *natural* jumps, in opposition to an intervention of the controller on the process giving rise to an impulse. In the latter case, at any time moment, we can apply an action from the set $\boldsymbol{A}^{\mathrm{i}}$ of impulsive actions to change instantly the state of the process according to a prescribed stochastic kernel Q on $\boldsymbol{X}$ given $\boldsymbol{X} \times \boldsymbol{A}^{\mathrm{i}}$. An intervention can lead not only to one single impulse but to any finite sequence of instantaneous impulses at the same time moment. As a result, the controlled process can take several different values at the same time moment, the intervention epoch. In the works of Hordijk and van der Duyn Schouten [11], [12], [13], [19], only one impulsive action at a time was allowed. As a consequence, the trajectory of the process was really a function of time, even if the intervention occurred immediately after a natural jump. In the works of Yushkevich [20], [21], [22], [23], the time scale was modified and split in order to make the trajectories a function of time. Therefore, a new theory of random processes had to be developed. In contrast, our aim is to use the standard theory of stochastic point processes [4], [15], [16], [17]. In this context, it is necessary to extend the state space to take into account the fact that the controlled system may have several different values at the same time moment. Our construction is based on a point process $(\Theta_n, Y_n)_{n\in\mathbb{N}}$, where Θ_n represents the sojourn time between two consecutive epochs induced either by a natural jump or by an intervention. Here Y_n is the new state vector of the form $(x_0, a_0, x_1, a_1, \dots, x_k, a_k, x_{k+1}, \Delta, \Delta, \dots)$, where x_0 corresponds to a possibly natural jump or to the value of the process just before the intervention. The pair (a_j, x_j) (for $j \geq 1$) indicates that the impulsive action a_j has been applied to the system, leading to a new location (jump) of the process denoted by x_j. The special impulsive action Δ means that the impulses have terminated and the artificial state Δ means the same. The space of all possible extended states is denoted by $\boldsymbol{Y}$ (this set will be precisely defined in the next section). The space of extended states resulting from interventions is denoted by $\boldsymbol{Y}^* = \boldsymbol{Y} \setminus \{(x_0, \Delta, \Delta, \dots),\ x_0 \in X\}$. Observe that $y = (x_0, \Delta, \Delta, \dots)$ means that no impulsive actions have been applied after a natural jump to state x_0.

We now present an informal description of the mechanism defining the controlled process $(\Theta_n, Y_n)_{n\in\mathbb{N}}$. In our framework, the interventions and gradual controls are determined through probability distributions on the appropriate spaces $\boldsymbol{Y}$ and $\boldsymbol{A}^{\mathrm{g}}$. The initial time moment 0 is very special. The initial state just before 0 is fixed and given by $Y_0 = (x_0, \Delta, \Delta, \dots)$, where x_0 is the initial location of the process. Moreover, the first sojourn time Θ_1 equals 0. Then the controller chooses a probability measure on $\boldsymbol{Y}$, generating the random variable Y_1 which is the next state immediately after time 0. After this initial procedure, the controlled process can be constructed iteratively. Having observed the state Y_n, the controller chooses the action u_n with the following components:

- a probability distribution on $\bar{\mathbb{R}}_+^*$ that generates the time of the (possible) next intervention which happens only in the case when no natural jumps occur earlier;
- a stochastic kernel on $\boldsymbol{A}^{\mathrm{g}}$ given $\mathbb{R}_+$ that describes the gradual control influencing the time of the next (possible) natural jump and its associated location;
- an intervention immediately after the natural jump, in the case when it happens before the planned intervention, that is, a probability distribution on $\boldsymbol{Y}$;
- a planned intervention, that is, a probability distribution on $\boldsymbol{Y}^*$. This last component is absent in the event that no interventions are allowed in the current state.

If the gradual action $a \in \boldsymbol{A}^{\mathrm{g}}$ is applied at the state $x \in X$ then the cost rate is $C^{\mathrm{g}}(x, a)$; any impulsive action $a \in \boldsymbol{A}^{\mathrm{i}}$ results in the immediate cost $c^{\mathrm{i}}(x, a)$. In the present paper we consider the discounted model on the infinite time horizon. Note that an intervention occurs at one time moment with a fixed value of the discounting coefficient so that it corresponds to a discrete-time MDP with a total expected cost.

Our model is closely related to those studied by Hordijk, van der Duyn Schouten and Yushkevich but presents important differences that we would like to emphasize. In particular, in [20]–[23] only nonrandomized gradual controls were considered. Moreover, in [11], [12],[13], and [19]– [23] the authors considered the times of intervention as stopping times with respect to the filtration generated by the controlled process. In our context the times of intervention were specified through probability distributions depending on the history of the process. In [5] and the references therein, the control strategies were past-history independent, deterministic, and several impulses at the same time moment were forbidden. Our framework is more general in the sense that we allow randomized policies. Moreover, we allow an instantaneous series of impulses which is not the case in [5], [11], [12], [13], and [19]. We would like to emphasize that [22] is the closest reference to our work because a series of impulses is allowed. The author studied the discounted cost control problem and showed that the value function is universally measurable and satisfies the Bellman equation. Moreover, the existence of a ε-optimal control strategy was proved.

When compared to the literature, our main contributions can be summarized as follows. Our main objective in this paper is to study the Bellman equation associated with this control problem, and to establish the existence of optimal and ε-optimal control strategies. We first show that, under some hypotheses, the continuous-time controlled process is nonexplosive. We provide sufficient conditions that ensure the existence and the uniqueness of a bounded measurable solution to the Bellman equation. It is proved that this solution can be calculated by successive iterations of the associated Bellman operator. Moreover, we show that the value function of our optimization problem satisfies this optimality equation. Two different sets of sufficient conditions are presented to ensure on the one hand the existence of an optimal control strategy, and on the other hand the existence of a ε-optimal control strategy. An interesting consequence of our previous results is as follows: the set of strategies that allow intervention at time $t = 0$ and only immediately after natural jumps is a sufficient set for the control problem under consideration.

The rest of the paper is organized as follows. Section 2 is devoted to the construction of CTMDPs on a general Borel state space $\boldsymbol{X}$ with both impulsive and continuous controls. In Section 3 we introduce the infinite-horizon performance criterion and several different classes of admissible strategies. Several preliminary results are also formulated here. The analysis of the Bellman equation and the existence of optimal and ε-optimal control strategies are discussed in Section 4.

2. The continuous-time Markov control process

The main goal of this section is to introduce the notation, as well as the parameters defining the model, and to present the construction of the controlled process. In particular, a measurable space $(\Omega, \mathcal{F})$ consisting of the canonical sample paths of the multivariate point process (Θ_n, Y_n) is introduced. Having defined the class of admissible strategies, we show the existence of a probability measure $\mathbb{P}^u_{x_0}$ with respect to which the controlled process (Θ_n, Y_n) has the required conditional distributions.

The following notation will be used in this paper: $\mathbb{N}$ denotes the set of natural numbers including 0, $\mathbb{N}^* = \mathbb{N} - \{0\}$, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_+$ denotes the set of nonnegative real numbers, $\mathbb{R}^*_+ = \mathbb{R}_+ - \{0\}$, $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$, and $\overline{\mathbb{R}}^*_+ = \mathbb{R}^*_+ \cup \{+\infty\}$. For any $q \in \mathbb{N}$, $\mathbb{N}_q$ is the set $\{0, 1, \dots, q\}$, and, for any $q \in \mathbb{N}^*$, $\mathbb{N}^*_q$ is the set $\{1, \dots, q\}$. The term *measure* will always refer to a countably additive, $\mathbb{R}_+$-valued set function. Let X be a Borel space, and denote by $\mathcal{B}(X)$ its associated Borel σ-algebra. For any set A, $\mathbf{1}_A$ denotes the indicator function of the set A. The set of measures defined on $(X, \mathcal{B}(X))$ is denoted by $\mathbb{M}(X)_+$, the set of probability measures defined on $(X, \mathcal{B}(X))$ is denoted by $\mathcal{P}(X)$, and $\mathcal{P}(X \mid Y)$ denotes the set of stochastic kernels on X given Y, where Y denotes a Borel space. For any point $x \in X$, δ_x denotes the Dirac measure defined by $\delta_x(\Gamma) = \mathbf{1}_\Gamma(x)$ for any $\Gamma \in \mathcal{B}(X)$. The set of bounded, real-valued measurable functions defined on X is denoted by $\mathbb{B}(X)$. Finally, the infimum over an empty set is understood to be equal to $+\infty$.

2.1. Parameters of the model

We will deal with a control model defined through the following elements.

- $\boldsymbol{X}$ is the state space, assumed to be a Borel space (i.e. a measurable subset of a complete and separable metric space).

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

- $\boldsymbol{A}$ is the action space, also assumed to be a Borel space. We denote by $\boldsymbol{A}^{\mathrm{i}} \in \mathcal{B}(\boldsymbol{A})$ and $\boldsymbol{A}^{\mathrm{g}} \in \mathcal{B}(\boldsymbol{A})$ the set of impulsive and gradual actions, respectively, satisfying $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{i}} \cup \boldsymbol{A}^{\mathrm{g}}$ with $\boldsymbol{A}^{\mathrm{i}} \cap \boldsymbol{A}^{\mathrm{g}} = \varnothing$.

- The set of feasible actions in state $x \in \boldsymbol{X}$ is $\boldsymbol{A}(x)$, which is a nonempty measurable subset of $\boldsymbol{A}$. Admissible impulsive and gradual actions in the state $x \in \boldsymbol{X}$ are denoted by $\boldsymbol{A}^{\mathrm{i}}(x) = \boldsymbol{A}(x) \cap \boldsymbol{A}^{\mathrm{i}}$ and $\boldsymbol{A}^{\mathrm{g}}(x) = \boldsymbol{A}(x) \cap \boldsymbol{A}^{\mathrm{g}}$. We suppose that $\mathbb{K}^{\mathrm{g}} = \{(x, a) \in \boldsymbol{X} \times \boldsymbol{A} \colon a \in \boldsymbol{A}^{\mathrm{g}}(x)\} \in \mathcal{B}(\boldsymbol{X} \times \boldsymbol{A}^{\mathrm{g}})$, with this set containing the graph of a measurable function from $\boldsymbol{X}$ to $\boldsymbol{A}^{\mathrm{g}}$ (necessarily $\boldsymbol{A}^{\mathrm{g}}(x) \neq \varnothing$ for all $x \in \boldsymbol{X}$), and that $\mathbb{K}^{\mathrm{i}} = \{(x, a) \in \boldsymbol{X} \times \boldsymbol{A}^{\mathrm{i}} \colon a \in \boldsymbol{A}^{\mathrm{i}}(x)\} \in \mathcal{B}(\mathbb{X}^{\mathrm{i}} \times \boldsymbol{A}^{\mathrm{i}})$, where $\mathbb{X}^{\mathrm{i}} = \{x \in \boldsymbol{X} \colon \boldsymbol{A}^{\mathrm{i}}(x) \neq \varnothing\} \in \mathcal{B}(\boldsymbol{X})$, with the set $\mathbb{K}^{\mathrm{i}}$ containing the graph of a measurable function from $\mathbb{X}^{\mathrm{i}}$ to $\boldsymbol{A}$.

- The stochastic kernel Q on $\boldsymbol{X}$ given $\mathbb{K}^{\mathrm{i}}$ describes the result of an impulsive action. In other words, if $x \in \mathbb{X}^{\mathrm{i}}$ and an impulsive action $a \in \boldsymbol{A}^{\mathrm{i}}(x)$ is applied, then the state of the process changes instantly according to the stochastic kernel Q.

- The signed kernel q on $\boldsymbol{X}$ given $\mathbb{K}^{\mathrm{g}}$ is the intensity of jumps governing the dynamic of the process between interventions. It satisfies $q(\boldsymbol{X} \mid x, a) = 0$ and $q(\Gamma \setminus \{x\} \mid x, a) \geq 0$ for any $(x, a) \in \mathbb{K}^{\mathrm{g}}$ and $\Gamma \in \mathcal{B}(\boldsymbol{X})$.

In our model, an intervention consists only of a finite sequence of pairs of impulsive action and associated jump. Actually, this finite sequence can be equivalently described by an infinite sequence of pairs of state and action, where the pairs are set to the fictitious action and state after a finite step. As a result, an intervention is an element of the set

$$\boldsymbol{Y} = \bigcup_{k \in \mathbb{N}} \boldsymbol{Y}_k \quad \text{with} \quad \boldsymbol{Y}_k = (\boldsymbol{X} \times \boldsymbol{A}^{\mathrm{i}})^k \times (\boldsymbol{X} \times \{\Delta\}) \times (\{\Delta\} \times \{\Delta\})^{\infty},$$

where Δ will play the role of the fictitious state and action. The dynamic of such sequences is governed by the Markov decision process (MDP) $\mathcal{M}^{\mathrm{i}}$ defined by $\mathcal{M}^{\mathrm{i}} = (\boldsymbol{X}_{\Delta}, \boldsymbol{A}^{\mathrm{i}}_{\Delta}, (\boldsymbol{A}^{\mathrm{i}}_{\Delta}(x))_{x \in \boldsymbol{X}_{\Delta}}, Q_{\Delta})$ where $\boldsymbol{X}_{\Delta}$, $\boldsymbol{A}^{\mathrm{i}}_{\Delta}$, and $(\boldsymbol{A}^{\mathrm{i}}_{\Delta}(x))_{x \in \boldsymbol{X}_{\Delta}}$ are the new state and action spaces augmented by the fictitious state Δ: $\boldsymbol{X}_{\Delta} = \boldsymbol{X} \cup \{\Delta\}$, $\boldsymbol{A}^{\mathrm{i}}_{\Delta} = \boldsymbol{A}^{\mathrm{i}} \cup \{\Delta\}$, and $\boldsymbol{A}^{\mathrm{i}}_{\Delta}(x) = \boldsymbol{A}^{\mathrm{i}}(x) \cup \{\Delta\}$ for $x \in \boldsymbol{X}$ and $\boldsymbol{A}^{\mathrm{i}}_{\Delta}(\Delta) = \{\Delta\}$. The dynamic is given by $Q_{\Delta}(\cdot \mid x, a) = Q(\cdot \mid x, a)$ for any $(x, a) \in \mathbb{K}^{\mathrm{i}}$ and $Q_{\Delta}(\{\Delta\} \mid x, a) = 1$ otherwise. For the model $\mathcal{M}^{\mathrm{i}}$, according to Ionescu Tulcea's theorem (see Proposition C.10 of [10]), there exists a unique strategic measure $\beta^b(\cdot \mid x)$ on $(\boldsymbol{X}_{\Delta} \times \boldsymbol{A}^{\mathrm{i}}_{\Delta})^{\infty}$ associated with the policy b and the initial distribution δ_x. Here and below, we use the standard terminology for MDPs: a policy is a sequence of past-dependent distributions on the action space; a Markov nonrandomized policy is a sequence $(\varphi^{\mathrm{i}}_j)_{j \in \mathbb{N}}$ of $\boldsymbol{A}^{\mathrm{i}}_{\Delta}$-valued mappings on $\boldsymbol{X}_{\Delta}$, and so on. Observe that β^b is in fact a stochastic kernel on $(\boldsymbol{X}_{\Delta} \times \boldsymbol{A}^{\mathrm{i}}_{\Delta})^{\infty}$ given $\boldsymbol{X}$; see Proposition C.10 of [10]. Since we only consider intervention as an element of $\boldsymbol{Y}$, we restrict the admissible policies to the set Ξ satisfying $\beta^b(\boldsymbol{Y} \mid x) = 1$ for $b \in \Xi$. In fact, we consider *randomized* interventions , consequently, an intervention is an element of

$$\mathcal{P}^{\boldsymbol{Y}} = \{\beta \in \mathcal{P}(\boldsymbol{Y} \mid \boldsymbol{X}) \colon \beta(\cdot \mid \cdot) = \beta^b(\cdot \mid \cdot) \text{ for some } b \in \Xi\},$$

and

$$\mathcal{P}^{\boldsymbol{Y}}(x) = \{\rho \in \mathcal{P}(\boldsymbol{Y}) \colon \rho(\cdot) = \beta^b(\cdot \mid x) \text{ for some } b \in \Xi\}$$

is the set of feasible interventions in state $x \in \boldsymbol{X}$. Observe that if an intervention is chosen in $\boldsymbol{Y}_0$, it actually means that the controller has not intervened on the process through impulsive actions. For technical reasons, it appears necessary to introduce the set $\boldsymbol{Y}^*$ of *real* interventions given by $\boldsymbol{Y}^* = \bigcup_{k=1}^{\infty} \boldsymbol{Y}_k$. The associated sets of real *randomized* interventions are defined by

$$\begin{aligned}\mathcal{P}^{\boldsymbol{Y}^*} = \{\beta \in \mathcal{P}(\boldsymbol{Y} \mid \boldsymbol{X}) \colon \beta(\cdot \mid \cdot) = \beta^b(\cdot \mid \cdot) \text{ for some } b \in \Xi \\ \text{and } \beta^b(\boldsymbol{Y}^* \mid x) = 1 \text{ for any } x \in \mathbb{X}^{\mathrm{i}}\}\end{aligned}$$

and

$$\mathcal{P}^{\boldsymbol{Y}^*}(x) = \{\rho \in \mathcal{P}(\boldsymbol{Y}) \colon \rho(\cdot) = \beta^b(\cdot \mid x) \text{ for some } b \in \Xi \text{ and } \beta^b(\boldsymbol{Y}^* \mid x) = 1\}$$

for $x \in \boldsymbol{X}$. Note that $\mathcal{P}^{\boldsymbol{Y}^*}(x) = \varnothing$ if $x \notin \mathbb{X}^{\mathrm{i}}$.

Finally, we end this subsection by introducing a *projection* mapping that will be used repeatedly in the paper. If $y \in \boldsymbol{Y}$ then there exists a unique $k \in \mathbb{N}$ such that $y \in \boldsymbol{Y}_k$. The $\boldsymbol{X}$-valued mapping $\bar{x}$ on $\boldsymbol{Y}$ is defined by

$$\bar{x}(y) = x_{k+1}.$$

2.2. Construction of the process

Having introduced the parameters of the model, we are now in position to construct the Markov controlled process. Let $\boldsymbol{Y}_\infty = \boldsymbol{Y} \cup \{y_\infty\}$ and $\Omega_n = \boldsymbol{Y} \times (\mathbb{R}_+^* \times \boldsymbol{Y})^n \times (\{\infty\} \times \{y_\infty\})^\infty$ for $n \in \mathbb{N}$. The canonical space Ω is defined as $\Omega = \bigcup_{n=1}^{\infty} \Omega_n \cup (\boldsymbol{Y} \times (\mathbb{R}_+^* \times \boldsymbol{Y})^\infty)$ and is endowed with its Borel σ-algebra denoted by $\mathcal{F}$. For notational convenience, $\omega \in \Omega$ will be represented as

$$\omega = (y_0, \theta_1, y_1, \theta_2, y_2, \ldots).$$

Here, $y_0 = (x_0, \Delta, \Delta, \ldots)$ is the initial state of the controlled point process ξ with values in $\boldsymbol{Y}$, defined below; $\theta_1 = 0$ and $y_1 \in \boldsymbol{Y}$ is the result of the initial intervention. The components $\theta_n > 0$ for $n \geq 2$ denote the sojourn times; y_n denotes the result of an intervention (if $y_n \in \boldsymbol{Y}^*$) or corresponds to a natural jump (if $y_n \in \boldsymbol{Y} \setminus \boldsymbol{Y}^*$). In the case in which $\theta_n < \infty$ and $\theta_{n+1} = \infty$, the trajectory has only n jumps and we put $y_m = y_\infty$ (artificial point) for all $m \geq n+1$.

The path up to $n \in \mathbb{N}$ is denoted by $h_n = (y_0, \theta_1, y_1, \theta_2, y_2, \ldots, \theta_n, y_n)$ and the collection of all such paths is denoted by $\boldsymbol{H}_n$. For $n \in \mathbb{N}$, introduce the mappings $Y_n \colon \Omega \to \boldsymbol{Y}_\infty$ by $Y_n(\omega) = y_n$ and, for $n \geq 2$, the mappings $\Theta_n \colon \Omega \to \overline{\mathbb{R}}_+^*$ by $\Theta_n(\omega) = \theta_n$; $\Theta_1(\omega) = 0$. The sequence $(T_n)_{n \in \mathbb{N}^*}$ of $\overline{\mathbb{R}}_+^*$-valued mappings is defined on Ω by $T_n(\omega) = \sum_{i=1}^n \Theta_i(\omega) = \sum_{i=1}^n \theta_i$ and $T_\infty(\omega) = \lim_{n\to\infty} T_n(\omega)$. For notational convenience, we denote by $H_n = (Y_0, \Theta_1, Y_1, \ldots, \Theta_n, Y_n)$ the n-term history process taking values in $\boldsymbol{H}_n$ for $n \in \mathbb{N}$.

The random measure μ associated with $(\Theta_n, Y_n)_{n \in \mathbb{N}}$ is a measure defined on $\mathbb{R}_+^* \times \boldsymbol{Y}$ by

$$\mu(\omega;\, \mathrm{d}t,\, \mathrm{d}y) = \sum_{n \geq 2} \mathbf{1}_{\{T_n(\omega) < \infty\}}\, \delta_{(T_n(\omega), Y_n(\omega))}(\mathrm{d}t,\, \mathrm{d}y).$$

For notational convenience, the dependence on ω will be ignored: instead of $\mu(\omega;\, \mathrm{d}t,\, \mathrm{d}y)$ we write $\mu(\mathrm{d}t,\, \mathrm{d}y)$. Define $\mathcal{F}_t = \sigma\{H_1\} \vee \sigma\{\mu(0, s] \times B) \colon s \leq t,\ B \in \mathcal{B}(\boldsymbol{Y})\}$ for $t \in \mathbb{R}_+$. Finally, we define the controlled process $\{\xi_t\}_{t \in \mathbb{R}_+}$ by

$$\xi_t(\omega) = \begin{cases} Y_n(\omega) & \text{if } T_n \leq t < T_{n+1} \text{ for } n \in \mathbb{N}^*, \\ y_\infty & \text{if } T_\infty \leq t, \end{cases}$$

and $\xi_{0-}(\omega) = Y_0 = y_0$ with $y_0 = (x_0, \Delta, \Delta, \ldots)$. Obviously, the controlled process $(\xi_t)_{t \in \mathbb{R}_+}$ can be equivalently described by the sequence $(\Theta_n, Y_n)_{n \in \mathbb{N}}$.

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

2.3. Admissible strategies and the conditional distribution of the controlled process

An admissible control strategy is a sequence $u = (u_n)_{n\in\mathbb{N}}$ such that $u_0 \in \mathcal{P}^{\boldsymbol{Y}}(x_0)$ and, for any $n \in \mathbb{N}^*$, u_n is given by

$$u_n = (\psi_n, \pi_n, \gamma_n^0, \gamma_n^1),$$

where ψ_n is a stochastic kernel on $\overline{\mathbb{R}}_+^*$ given $\boldsymbol{H}_n$ satisfying $\psi_n(\cdot \mid h_n) = \delta_{+\infty}(\cdot)$ for any $h_n = (y_0, \theta_1, \dots, \theta_n, y_n) \in \boldsymbol{H}_n$ with $\overline{x}(y_n) \notin \mathbb{X}^{\mathrm{i}}$, π_n is a stochastic kernel on $\boldsymbol{A}^{\mathrm{g}}$ given $\boldsymbol{H}_n \times \mathbb{R}_+^*$ satisfying $\pi_n(\boldsymbol{A}^{\mathrm{g}}(\overline{x}(y_n)) \mid h_n, t) = 1$ for any $t \in \mathbb{R}_+^*$, with $h_n = (y_0, \theta_1, \dots, \theta_n, y_n) \in \boldsymbol{H}_n$, γ_n^0 is a stochastic kernel on $\boldsymbol{Y}$ given $\boldsymbol{H}_n \times \mathbb{R}_+^* \times \boldsymbol{X}$ satisfying $\gamma_n^0(\cdot \mid h_n, t, \cdot) \in \mathcal{P}^{\boldsymbol{Y}}$ for any $h_n \in \boldsymbol{H}_n$ and $t \in \mathbb{R}_+^*$, and γ_n^1 is a stochastic kernel on $\boldsymbol{Y}$ given $\boldsymbol{H}_n$ satisfying $\gamma_n^1(\cdot \mid h_n) \in \mathcal{P}^{\boldsymbol{Y}^*}(\overline{x}(y_n))$ for any $h_n = (y_0, \theta_1, \dots, \theta_n, y_n) \in \boldsymbol{H}_n$ with $\overline{x}(y_n) \in \mathbb{X}^{\mathrm{i}}$; if $\overline{x}(y_n) \notin \mathbb{X}^{\mathrm{i}}$ then $\gamma_n^1(\cdot \mid h_n) = \delta_{(\overline{x}(y_n),\Delta,\Delta,\dots)}(\cdot)$.

The above conditions apply when $y_n \neq y_\infty$; otherwise, all the values of $\psi_n(\cdot \mid h_n)$, $\pi_n(\cdot \mid h_n, t)$, $\gamma_n^0(\cdot \mid h_n, t, \cdot)$, and $\gamma_n^1(\cdot \mid h_n)$ may be arbitrary.

The set of admissible control strategies is denoted by $\mathcal{U}$. In what follows, we use the notation $\gamma_n = (\gamma_n^0, \gamma_n^1)$.

Suppose that a strategy $u = (u_n)_{n\in\mathbb{N}} \in \mathcal{U}$ is fixed with $u_n = (\psi_n, \pi_n, \gamma_n^0, \gamma_n^1)$ for $n \in \mathbb{N}^*$. We introduce the intensity of the natural jumps as

$$\lambda_n(\Gamma_x, h_n, t) = \int_{\boldsymbol{A}^{\mathrm{g}}} \overline{q}(\Gamma_x \mid \overline{x}(y_n), a)\pi_n(\mathrm{d}a \mid h_n, t),$$

where $\overline{q}(\Gamma_x \mid x, a) = q(\Gamma_x \setminus \{x\} \mid x, a)$ for $(x, a) \in \boldsymbol{X} \times \boldsymbol{A}^{\mathrm{g}}$, and the rate of the natural jumps as

$$\Lambda_n(\Gamma_x, h_n, t) = \int_{(0,t]} \lambda_n(\Gamma_x, h_n, s)\,\mathrm{d}s$$

for any $n \in \mathbb{N}^*$, $\Gamma_x \in \mathcal{B}(\boldsymbol{X})$, and $h_n = (y_0, \theta_1, y_1, \dots, \theta_n, y_n) \in \boldsymbol{H}_n$. Now, for any $n \in \mathbb{N}^*$, the stochastic kernel G_n on $\boldsymbol{Y}_\infty \times \overline{\mathbb{R}}_+^*$ given $\boldsymbol{H}_n$ is defined by

$$G_n(\{+\infty\} \times \{y_\infty\} \mid h_n) = \delta_{y_n}(\{y_\infty\}) + \delta_{y_n}(\boldsymbol{Y})\mathrm{e}^{-\Lambda_n(\boldsymbol{X},h_n,+\infty)}\psi_n(\{+\infty\} \mid h_n)$$

and

$$\begin{aligned} G_n&(\Gamma_\Theta \times \Gamma_y \mid h_n) \\ &= \delta_{y_n}(\boldsymbol{Y})\bigg[\gamma_n^1(\Gamma_y \mid h_n)\int_{\Gamma_\theta} \mathrm{e}^{-\Lambda_n(\boldsymbol{X},h_n,t)}\psi_n(\mathrm{d}t \mid h_n) \\ &\qquad + \int_{\Gamma_\theta}\int_{\boldsymbol{X}} \psi_n([t,\infty] \mid h_n)\gamma_n^0(\Gamma_y \mid h_n, t, x)\lambda_n(\mathrm{d}x, h_n, t)\mathrm{e}^{-\Lambda_n(\boldsymbol{X},h_n,t)}\,\mathrm{d}t\bigg], \end{aligned} \tag{2.1}$$

where $\Gamma_y \in \mathcal{B}(\boldsymbol{Y})$, $\Gamma_\Theta \in \mathcal{B}(\mathbb{R}_+^*)$, and $h_n = (y_0, \theta_1, y_1, \dots, \theta_n, y_n) \in \boldsymbol{H}_n$. Note that the kernel γ_n^1 does not appear in the equation for G_n if $\overline{x}(y_n) \notin \mathbb{X}^{\mathrm{i}}$.

Consider an admissible strategy $u \in \mathcal{U}$ and an initial state $x_0 \in \boldsymbol{X}$. From Theorem 3.6 of [15] (or Remark 3.43 of [16]), there exists a probability $\mathbb{P}_{x_0}^u$ on $(\Omega, \mathcal{F})$ such that the restriction of $\mathbb{P}_{x_0}^u$ to $(\Omega, \mathcal{F}_0)$ is given by

$$\mathbb{P}_{x_0}^u(\{Y_0\} \times \{0\} \times \Gamma_y \times (\overline{\mathbb{R}}_+^* \times \boldsymbol{Y}_\infty)^\infty) = u_0(\Gamma_y \mid x_0) \tag{2.2}$$

for any $\Gamma_y \in \mathcal{B}(\boldsymbol{Y})$, and the positive random measure ν defined on $\mathbb{R}_+^* \times \boldsymbol{Y}$ by

$$\nu(\mathrm{d}t,\ \mathrm{d}y) = \sum_{n\in\mathbb{N}^*} \frac{G_n(\mathrm{d}t - T_n,\ \mathrm{d}y \mid H_n)}{G_n([t - T_n, +\infty] \times \boldsymbol{Y}_\infty \mid H_n)} \mathbf{1}_{\{T_n < t \le T_{n+1}\}} \tag{2.3}$$

is the predictable projection of μ with respect to $\mathbb{P}^u_{x_0}$.

Remark 2.1. Observe that $\mathcal{F}_{T_n}$ is the σ-algebra generated by the random variable H_n for $n \in \mathbb{N}^*$. The conditional distribution of (Y_{n+1}, Θ_{n+1}) given $\mathcal{F}_{T_n}$ under $\mathbb{P}^u_{x_0}$ is determined by $G_n(\cdot \mid H_n)$, and the conditional survival function of Θ_{n+1} given $\mathcal{F}_{T_n}$ under $\mathbb{P}^u_{x_0}$ is given by $G_n([t, +\infty] \times \boldsymbol{Y}_\infty \mid H_n)$.

3. Optimization problem and preliminary results

The objective of this section is to introduce the infinite-horizon performance criterion we are concerned with and several different classes of admissible strategies. Some preliminary results are established. In particular, assuming that the process is nonexplosive, a discounted version of the so-called Dynkin formula associated with the controlled process is derived (see Lemma 3.2).

The first result provides a decomposition of the predictable projection ν of the process into two parts: one part being related to the component $(\gamma_n^0)_{n\in\mathbb{N}^*}$ of an admissible control strategy and the other part being related to the component $(\gamma_n^1)_{n\in\mathbb{N}^*}$

Lemma 3.1. *The predictable projection of the random measure μ is given by $\nu = \nu_0 + \nu_1$ with*

$$\nu_0(\mathrm{d}s,\ \mathrm{d}y) = \int_{\boldsymbol{A}^{\mathrm{g}}} \int_{\boldsymbol{X}} \gamma^0(\mathrm{d}y \mid x, s) \overline{q}(\mathrm{d}x \mid \overline{x}(\xi_{s-}), a) \pi(\mathrm{d}a \mid s)\, \mathrm{d}s,$$

$$\nu_1(\mathrm{d}s,\ \mathrm{d}y) = \sum_{n\in\mathbb{N}^*} \gamma_n^1(\mathrm{d}y \mid H_n)\, \mathbf{1}_{\{T_n < s \le T_{n+1}\}} \frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)},$$

and

$$\begin{aligned}\gamma^0(\mathrm{d}y \mid x, t) &= \sum_{n\in\mathbb{N}^*} \mathbf{1}_{\{T_n < t \le T_{n+1}\}}\, \gamma_n^0(\mathrm{d}y \mid H_n, t - T_n, x), \pi(\mathrm{d}a \mid t) \\ &= \sum_{n\in\mathbb{N}^*} \mathbf{1}_{\{T_n < t \le T_{n+1}\}}\, \pi_n(\mathrm{d}a \mid H_n, t - T_n)\end{aligned}$$

for $t \in \mathbb{R}_+^*$.

Proof. First observe that by using integration by parts we obtain

$$G_n([t, +\infty] \times \boldsymbol{Y}_\infty \mid h_n) = \delta_{y_n}(\{y_\infty\}) + \delta_{y_n}(\boldsymbol{Y}) \mathrm{e}^{-\Lambda_n(\boldsymbol{X}, h_n, t)} \psi_n([t, +\infty] \mid h_n).$$

Now, recalling the definition of ν (see (2.3)) in terms of G (see (2.1)), a straightforward calculation gives the result.

The cost rate C^{g} associated with a gradual action is a real-valued mapping defined on $\mathbb{K}^{\mathrm{g}}$. The cost associated with an intervention $y = (x_0, a_0, x_1, a_1, \ldots) \in \boldsymbol{Y}$ is given by $C^{\mathrm{i}}(y) = \sum_{k\in\mathbb{N}} c^{\mathrm{i}}(x_k, a_k)$, where c^{i} is a real-valued mapping defined on $\boldsymbol{X}_\Delta \times \boldsymbol{A}^{\mathrm{i}}_\Delta$ satisfying $c^{\mathrm{i}}(x, a) = 0$ if $(x, a) \notin \mathbb{K}^{\mathrm{i}}$. For any $(x, a) \in \mathbb{K}^{\mathrm{i}}$, $c^{\mathrm{i}}(x, a)$ corresponds to the cost associated with a single jump at $x \in \boldsymbol{X}$ resulting from the impulsive action $a \in \boldsymbol{A}^{\mathrm{i}}(x)$. The cost associated with a

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

randomized intervention $\beta \in \mathcal{P}^{Y}(x)$ for $x \in \boldsymbol{X}$ is given by $\int_{\boldsymbol{Y}} C^{\mathrm{i}}(y)\beta(\mathrm{d}y \mid x)$. Therefore, the infinite-horizon discounted performance criterion corresponding to an admissible control strategy $u \in \mathcal{U}$ is defined by

$$\mathcal{V}(u, x_0) = \int_{\boldsymbol{Y}} C^{\mathrm{i}}(y)u_0(\mathrm{d}y \mid x_0) + \mathbb{E}^{u}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s} \int_{\boldsymbol{A}^{\mathrm{g}}} C^{\mathrm{g}}(\overline{x}(\xi_{s-}), a)\pi(\mathrm{d}a \mid s)\,\mathrm{d}s\bigg] + \mathbb{E}^{u}_{x_0}\bigg[\int_{(0,\infty)\times \boldsymbol{Y}} \mathrm{e}^{-\eta s} C^{\mathrm{i}}(y)\mu(\mathrm{d}s,\,\mathrm{d}y)\bigg]. \tag{3.1}$$

In the previous expression, where $\eta > 0$ is the discount factor, $\mathcal{V}(u, x_0)$ is understood to be equal to $+\infty$ if the integrals of both the positive and negative parts of the integrand are infinite. Note that, for any control strategy $u \in \mathcal{U}$, the function $\mathcal{V}(u, \cdot)$ is measurable. The optimization problem under consideration is to minimize $\mathcal{V}(u, x_0)$ within the class of admissible strategies $u \in \mathcal{U}$, where x_0 is the initial state. In the following a control strategy $u \in \mathcal{U}$ is called

- *nonrandomized stationary*, if $\psi_n(\cdot \mid h_n) = \delta_{\psi^s(\overline{x}(y_n))}(\cdot)$, $\pi_n(\cdot \mid h_n, t) = \delta_{\varphi^s(\overline{x}(y_n))}(\cdot)$, $\gamma^0_n(\cdot \mid h_n, t, \cdot) = \beta^{b_0}(\cdot \mid \cdot)$, and $\gamma^1_n(\cdot \mid h_n) = \beta^{b_1}(\cdot \mid \overline{x}(y_n))$, where ψ^s and φ^s are measurable maps from $\boldsymbol{X}$ to $\overline{\mathbb{R}}^*_+$ and $\boldsymbol{X}$ to $\boldsymbol{A}^{\mathrm{g}}$, respectively, and b_0 and b_1 are nonrandomized stationary policies in $\mathcal{M}^{\mathrm{i}}$;
- *nonrandomized almost stationary* when b_0 and b_1 in the above definition are Markov nonrandomized policies;
- *uniformly or persistently* optimal (respectively ε-optimal for $\varepsilon > 0$), if $\mathcal{V}(u, x_0) = \inf_{v\in\mathcal{U}} \mathcal{V}(v, x_0)$ (respectively $\mathcal{V}(u, x_0) \le \mathcal{V}(v, x_0) + \varepsilon$ for any $v \in \mathcal{U}$) simultaneously for all $x_0 \in \boldsymbol{X}$ and, hence, for any initial distribution.

The following lemma provides a discounted version of the so-called Dynkin formula associated with the controlled process $(\xi_t)_{t\in\mathbb{R}_+}$

Lemma 3.2. *Suppose that a strategy* $u = (u_n)_{n\in\mathbb{N}} \in \mathcal{U}$ *is fixed with* $u_n = (\psi_n, \pi_n, \gamma^0_n, \gamma^1_n)$ *for* $n \in \mathbb{N}^*$ *satisfying* $\mathbb{P}^u_{x_0}(T_\infty = +\infty) = 1$. *Then we have*

$$\begin{aligned}
&\mathbb{E}^u_{x_0}[\mathrm{e}^{-\eta t} W(\overline{x}(\xi_t))] \\
&\quad = \int_{\boldsymbol{Y}} W(\overline{x}(y))u_0(\mathrm{d}y \mid x_0) \\
&\qquad + \mathbb{E}^u_{x_0}\bigg[\int_0^t \int_{\boldsymbol{A}^{\mathrm{g}}} \mathrm{e}^{-\eta s}\bigg[-\eta W(\overline{x}(\xi_s)) + \int_{\boldsymbol{X}}\int_{\boldsymbol{Y}} W(\overline{x}(y))\gamma^0(\mathrm{d}y \mid x, s)\overline{q}(\mathrm{d}x \mid \overline{x}(\xi_s), a) \\
&\qquad\qquad - W(\overline{x}(\xi_s))\overline{q}(\boldsymbol{X} \mid \overline{x}(\xi_s), a)\bigg]\pi(\mathrm{d}a \mid s)\,\mathrm{d}s\bigg] \\
&\qquad + \mathbb{E}^u_{x_0}\bigg[\sum_{n\in\mathbb{N}^*}\int_{(T_n\wedge t, T_{n+1}\wedge t]}\int_{\boldsymbol{Y}} \mathrm{e}^{-\eta s}[W(\overline{x}(y)) - W(\overline{x}(\xi_{s-}))] \\
&\qquad\qquad \times \gamma^1_n(\mathrm{d}y \mid H_n)\frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)}\bigg],
\end{aligned}$$

for any bounded, real-valued measurable function W *defined on* $\boldsymbol{X}$.

Proof. Recalling that ν is the predictable projection of μ and that $\mathbb{P}^u_{x_0}(T_\infty = +\infty) = 1$, it follows by using the product formula for functions of bounded variation (see, e.g. Theorem A.4.6

of [17]) that

$$
\begin{aligned}
&\mathbb{E}^u_{x_0}[\mathrm{e}^{-\eta t} W(\overline{x}(\xi_t))] \\
&\quad = \mathbb{E}^u_{x_0}[W(\overline{x}(y_1))] - \mathbb{E}^u_{x_0}\biggl[\int_0^t \eta \mathrm{e}^{-\eta s} W(\overline{x}(\xi_s))\,\mathrm{d}s\biggr] \\
&\qquad + \mathbb{E}^u_{x_0}\biggl[\int_{(0,t]}\int_{\boldsymbol{Y}} \mathrm{e}^{-\eta s}[W(\overline{x}(y)) - W(\overline{x}(\xi_{s-}))]\gamma^0(\mathrm{d}y \mid x, s) \\
&\qquad\qquad \times \int_{\boldsymbol{A}^{\mathrm{g}}}\int_{\boldsymbol{X}} \overline{q}(\mathrm{d}x \mid \overline{x}(\xi_{s-}), a)\pi(\mathrm{d}a \mid s)\,\mathrm{d}s\biggr] \\
&\qquad + \mathbb{E}^u_{x_0}\biggl[\sum_{n\in\mathbb{N}^*}\int_{(T_n\wedge t, T_{n+1}\wedge t]}\int_{\boldsymbol{Y}} \mathrm{e}^{-\eta s}[W(\overline{x}(y)) - W(\overline{x}(\xi_{s-}))] \\
&\qquad\qquad\qquad \times \gamma^1_n(\mathrm{d}y \mid H_n)\frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)}\biggr].
\end{aligned}
$$

Now, from (2.2), it follows that $\mathbb{E}^u_{x_0}[W(\overline{x}(y_1))] = \int_{\boldsymbol{Y}} W(\overline{x}(y))u_0(\mathrm{d}y \mid x_0)$, showing the result.

4. Main results

This section is devoted to the analysis of the so-called Bellman equation associated with the control problem described in the previous section, and to the existence of optimal and ε-optimal control strategies. The first result (Proposition 4.1) ensures that the continuous-time controlled process is nonexplosive under some hypotheses. Then we provide two different sets of conditions (Assumption (C.1) and (C.2)) ensuring the existence of a bounded measurable solution to the Bellman equation. More precisely, in Propositions 4.2 and 4.3 we prove that this solution can be calculated by the successive iteration of the associated Bellman operator, leading either to an upper-semicontinuous or to a lower-semicontinuous solution. Moreover, we show in Theorem 4.1 and Corollary 4.1, on the one hand, the existence of an optimal control strategy and, on the other hand, the existence of a ε-optimal control strategy. We also prove that the value function of the optimization problem under consideration satisfies this optimality equation and, as a consequence, the bounded solution of the Bellman equation is unique. We exhibit the decomposition of the state space into two disjoint subsets $\boldsymbol{X}^{\mathrm{i}}$ and $\boldsymbol{X}^{\mathrm{g}}$, where, roughly speaking, one should apply a gradual action if the current state is in $\boldsymbol{X}^{\mathrm{g}}$, and an impulsive action if the current state is in $\boldsymbol{X}^{\mathrm{i}}$, to obtain an optimal or a ε-optimal strategy, depending on the assumptions under consideration. Another important and interesting consequence of our previous results is as follows: the set of strategies that allow intervention at time $t = 0$ and only immediately after natural jumps is a sufficient set for the control problem under study. (See Theorem 4.1 and Corollary 4.1.)

The Bellman equation reads as follows:

$$
\begin{aligned}
&\inf_{a\in\boldsymbol{A}^{\mathrm{g}}(x)}\biggl\{-\eta V(x) + \int_{\boldsymbol{X}} V(\tilde{x})\overline{q}(\mathrm{d}\tilde{x} \mid x, a) - V(x)\overline{q}(\boldsymbol{X} \mid x, a) + C^{\mathrm{g}}(x, a)\biggr\} \\
&\wedge \inf_{a\in\boldsymbol{A}^{\mathrm{i}}(x)}\biggl\{-V(x) + \int_{\boldsymbol{X}} V(\tilde{x})Q(\mathrm{d}\tilde{x} \mid x, a) + c^{\mathrm{i}}(x, a)\biggr\} \\
&\quad = 0
\end{aligned}
\tag{4.1}
$$

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

for any $x \in \boldsymbol{X}$. If V is a solution to (4.1), we introduce the following subsets of $\boldsymbol{X}$:

$$\boldsymbol{X}^{\mathrm{g}} = \left\{ x \in \boldsymbol{X} \colon \eta V(x) = \inf_{a \in \boldsymbol{A}^{\mathrm{g}}(x)} \left\{ \int_{\boldsymbol{X}} V(\tilde{x}) \overline{q}(\mathrm{d}\tilde{x} \mid x, a) - V(x) \overline{q}(\boldsymbol{X} \mid x, a) + C^{\mathrm{g}}(x, a) \right\} \right\},$$

and

$$\boldsymbol{X}^{\mathrm{i}} = \boldsymbol{X} \setminus \boldsymbol{X}^{\mathrm{g}} \subset \left\{ x \in \boldsymbol{X} \colon V(x) = \inf_{a \in \boldsymbol{A}^{\mathrm{i}}(x)} \left\{ \int_{\boldsymbol{X}} V(\tilde{x}) Q(\mathrm{d}\tilde{x} \mid x, a) + c^{\mathrm{i}}(x, a) \right\} \right\}.$$

These sets will be used to construct an optimal or a ε-optimal strategy in the proof of Theorem 4.1. Below, we provide conditions under which there exists a measurable bounded solution to the Bellman equation. These conditions also guarantee that the sets $\boldsymbol{X}^{\mathrm{g}}$ and $\boldsymbol{X}^{\mathrm{i}}$ are measurable.

Assumption A. *There exists a constant $K \in \mathbb{R}$ such that, for any $x \in \boldsymbol{X}$, $a^{\mathrm{g}} \in \boldsymbol{A}^{\mathrm{g}}(x)$, and $a^{\mathrm{i}} \in \boldsymbol{A}^{\mathrm{i}}(x)$,*

(A.1) $\overline{q}(\boldsymbol{X} \mid x, a^{\mathrm{g}}) \leq K$,

(A.2) $|C^{\mathrm{g}}(x, a^{\mathrm{g}})| \leq K$,

(A.3) $c^{\mathrm{i}}(x, a^{\mathrm{i}}) \geq 0$.

Assumption B. *There exists $\underline{c} > 0$ such that $c^{\mathrm{i}}(x, a) \geq \underline{c}$ for any $(x, a) \in \mathbb{K}^{\mathrm{i}}$.*

The following proposition gives a sufficient condition for nonexplosion.

Proposition 4.1. *Suppose that Assumptions A and B hold. If $u \in \mathcal{U}$ and $\mathcal{V}(u, x_0) < \infty$, then $\mathbb{P}^u_{x_0}(T_\infty < \infty) = 0$.*

Proof. From Assumption A and the definition of the cost (3.1), we have

$$\begin{aligned} \mathcal{V}(u, x_0) &\geq -\frac{K}{\eta} + \mathbb{E}^u_{x_0}\left[\int_{(0,\infty[\times \boldsymbol{Y}} \mathrm{e}^{-\eta s} C^{\mathrm{i}}(y) \mu(\mathrm{d}s,\ \mathrm{d}y) \right] \\ &\geq -\frac{K}{\eta} + \mathbb{E}^u_{x_0}\left[\sum_{n \in \mathbb{N}^*} \int_{(T_n, T_{n+1}] \times \boldsymbol{Y}} \mathrm{e}^{-\eta s} C^{\mathrm{i}}(y) \gamma^1_n(\mathrm{d}y \mid H_n) \frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)} \right]. \end{aligned}$$

Now, observe that if $\overline{x}(Y_n) \notin \mathbb{X}^{\mathrm{i}}$ then the measure

$$\mathrm{e}^{-\eta s} \gamma^1_n(\mathrm{d}y \mid H_n) \frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)}$$

is 0 on the set $(T_n, T_{n+1}] \times \boldsymbol{Y}$ and if $\overline{x}(Y_n) \in \mathbb{X}^{\mathrm{i}}$ then $\gamma^1_n(\cdot \mid H_n) \in \mathcal{P}^{\boldsymbol{Y}^*}(\overline{x}(Y_n))$, and that $C^{\mathrm{i}}(y) \geq \underline{c}$ for any $y \in \boldsymbol{Y}^*$ by Assumption B. Consequently,

$$\mathcal{V}(u, x_0) \geq -\frac{K}{\eta} + \underline{c}\mathbb{E}^u_{x_0}\left[\sum_{n \in \mathbb{N}^*} \int_{(T_n, T_{n+1}] \times \boldsymbol{Y}} \mathrm{e}^{-\eta s} \gamma^1_n(\mathrm{d}y \mid H_n) \frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)} \right]. \tag{4.2}$$

Moreover, from Assumption (A.1) we obtain

$$\mathbb{E}^u_{x_0}\left[\int_0^{+\infty} \int_{\boldsymbol{A}^{\mathrm{g}}} \mathrm{e}^{-\eta s} \int_{\boldsymbol{X}} \int_{\boldsymbol{Y}} \gamma^0(\mathrm{d}y \mid x, s) \overline{q}(\mathrm{d}x \mid \overline{x}(\xi_s), a) \pi(\mathrm{d}a \mid s)\, \mathrm{d}s \right] \leq \frac{K}{\eta}. \tag{4.3}$$

Combining (4.2) and (4.3), we have

$$\mathbb{E}^u_{x_0}\bigg[\int_0^{+\infty}\int_{\boldsymbol{Y}} \mathrm{e}^{-\eta s}\mu(\mathrm{d}s,\ \mathrm{d}z)\bigg] = \mathbb{E}^u_{x_0}\bigg[\int_0^{+\infty}\int_{\boldsymbol{Y}} \mathrm{e}^{-\eta s}\nu(\mathrm{d}s,\ \mathrm{d}z)\bigg] \leq \frac{1}{\underline{c}}\bigg[\mathcal{V}(u, x_0) + \frac{K}{\eta}\bigg] + \frac{K}{\eta}. \tag{4.4}$$

However, if $\mathbb{P}^u_{x_0}(T_\infty < \infty) > 0$ then

$$\mathbb{E}^u_{x_0}\bigg[\int_0^{+\infty}\int_{\boldsymbol{Y}} \mathrm{e}^{-\eta s}\mu(\mathrm{d}s,\ \mathrm{d}z)\bigg] \geq \mathbb{E}^u_{x_0}[\mathrm{e}^{-\eta T_\infty}\mu(\mathbb{R}^*_+, \boldsymbol{Y})\,\mathbf{1}_{\{T_\infty<\infty\}}] = +\infty. \tag{4.5}$$

From (4.4) and (4.5), it follows that if $u \in \mathcal{U}$ satisfies $\mathcal{V}(u, x_0) < \infty$ then $\mathbb{P}^u_{x_0}(T_\infty < \infty) = 0$, showing the result.

In Assumption C below, we assume that metrizable topologies in the spaces $\boldsymbol{X}$ and $\boldsymbol{A}$ are fixed.

Assumption C. (C.1) *The sets $\mathbb{K}^{\mathrm{g}}$ and $\mathbb{K}^{\mathrm{i}}$ are open in $\boldsymbol{X} \times \boldsymbol{A}^{\mathrm{g}}$ and $\mathbb{X}^{\mathrm{i}} \times \boldsymbol{A}^{\mathrm{i}}$, respectively. For any continuous bounded function F on $\boldsymbol{X}$, the functions $\int_{\boldsymbol{X}} F(z)\overline{q}(\mathrm{d}z \mid x, a)$ and $\int_{\boldsymbol{X}} F(z)Q(\mathrm{d}z \mid x, a)$ are continuous on $\mathbb{K}^{\mathrm{g}}$ and $\mathbb{K}^{\mathrm{i}}$, respectively. The functions C^{g} and c^{i} are upper semicontinuous on $\mathbb{K}^{\mathrm{g}}$ and $\mathbb{K}^{\mathrm{i}}$, respectively.*

(C.2) *The sets $\boldsymbol{A}^{\mathrm{g}}$ and $\boldsymbol{A}^{\mathrm{i}}$ are compact, and the sets $\mathbb{K}^{\mathrm{g}}$ and $\mathbb{K}^{\mathrm{i}}$ are closed in $\boldsymbol{X} \times \boldsymbol{A}^{\mathrm{g}}$ and $\mathbb{X}^{\mathrm{i}} \times \boldsymbol{A}^{\mathrm{i}}$, respectively. For any continuous bounded function F on $\mathbf{X}$, the functions $\int_{\boldsymbol{X}} F(z)\overline{q}(\mathrm{d}z \mid x, a)$ and $\int_{\boldsymbol{X}} F(z)Q(\mathrm{d}z \mid x, a)$ are continuous on $\mathbb{K}^{\mathrm{g}}$ and $\mathbb{K}^{\mathrm{i}}$, respectively. The functions C^{g} and c^{i} are lower semicontinuous on $\mathbb{K}^{\mathrm{g}}$ and $\mathbb{K}^{\mathrm{i}}$, respectively.*

Introduce the stochastic kernel $\widetilde{P}$ on $\boldsymbol{X}$ given $\mathbb{K}^{\mathrm{g}}$ $\widetilde{P}(\Gamma \mid x, a) = (1/K)[\overline{q}(\Gamma \mid x, a) + \delta_x(\Gamma)[K - \overline{q}(\boldsymbol{X} \mid x, a)]]$ for any $\Gamma \in \mathcal{B}(\boldsymbol{X})$ and $(x, a) \in \mathbb{K}^{\mathrm{g}}$, and consider the mapping $\mathfrak{B}$ defined on $\mathbb{B}(\boldsymbol{X})$ by

$$\begin{aligned} \mathfrak{B}F(x) = &\inf_{a\in \boldsymbol{A}^{\mathrm{g}}(x)}\bigg\{\frac{K}{K+\eta}\int_{\boldsymbol{X}} F(\widetilde{x})\widetilde{P}(\mathrm{d}\widetilde{x} \mid x, a) + \frac{1}{K+\eta}C^{\mathrm{g}}(x, a)\bigg\} \\ &\wedge \inf_{a\in \boldsymbol{A}^{\mathrm{i}}(x)}\bigg\{\int_{\boldsymbol{X}} F(\widetilde{x})Q(\mathrm{d}\widetilde{x} \mid x, a) + c^{\mathrm{i}}(x, a)\bigg\} \end{aligned} \tag{4.6}$$

for any $F \in \mathbb{B}(\boldsymbol{X})$. The mapping $\mathfrak{B}$ will be called the *Bellman operator* for further reference.

The next two propositions ensure, under two different sets of conditions, the existence of an upper-semicontinuous or a lower-semicontinuous solution of the Bellman equation, the measurability of the corresponding sets $\boldsymbol{X}^{\mathrm{g}}$ and $\boldsymbol{X}^{\mathrm{i}}$, and the existence of Borel-measurable mappings $\varphi^{\mathrm{i}}\colon \boldsymbol{X}^{\mathrm{i}} \to \boldsymbol{A}^{\mathrm{i}}$ and $\varphi^{\mathrm{g}}\colon \boldsymbol{X}^{\mathrm{g}} \to \boldsymbol{A}^{\mathrm{g}}$ that will be used to construct optimal strategies.

Proposition 4.2. *Suppose that Assumptions A and (C.1) hold. Then the decreasing sequence of functions $(V_i)_{i\in\mathbb{N}}$ defined iteratively by $V_{i+1} = \mathfrak{B}V_i$ with $V_0 = K/\eta$ belongs to $\mathbb{B}(\boldsymbol{X})$ and converges to a bounded upper-semicontinuous function V on $\boldsymbol{X}$ satisfying the Bellman equation (4.1). Moreover, the corresponding sets $\boldsymbol{X}^{\mathrm{g}}$ and $\boldsymbol{X}^{\mathrm{i}}$ are measurable and, for any $\varepsilon > 0$, there exist Borel-measurable mappings $\varphi^{\mathrm{i}}\colon \boldsymbol{X}^{\mathrm{i}} \to \boldsymbol{A}^{\mathrm{i}}$ and $\varphi^{\mathrm{g}}\colon \boldsymbol{X}^{\mathrm{g}} \to \boldsymbol{A}^{\mathrm{g}}$, such that*

$$\varphi^{\mathrm{i}}(z) \in \bigg\{a \in \boldsymbol{A}^{\mathrm{i}}(z)\colon \int_{\boldsymbol{X}} V(\widetilde{x})Q(\mathrm{d}\widetilde{x} \mid z, a) + c^{\mathrm{i}}(z, a) \leq V(z) + \varepsilon\bigg\}, \tag{4.7}$$

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

for any $z \in \boldsymbol{X}^{\mathrm{i}}$ *and*

$$\varphi^{\mathrm{g}}(z) \in \left\{a \in \boldsymbol{A}^{\mathrm{g}}(z) \colon \int_{\boldsymbol{X}} V(\widetilde{x})\overline{q}(\mathrm{d}\widetilde{x} \mid z, a) - V(z)\overline{q}(\boldsymbol{X} \mid z, a) + C^{\mathrm{g}}(z, a) \leq \eta V(z) + \varepsilon\right\} \tag{4.8}$$

for any $z \in \boldsymbol{X}^{\mathrm{g}}$.

Proof. By using simple algebraic manipulations and (A.1)–(A.2), it is easy to show that $V \in \mathbb{B}(\boldsymbol{X})$ is a solution of the Bellman equation (4.1) if and only if $V \in \mathbb{B}(\boldsymbol{X})$ and $V = \mathfrak{B}V$ holds. Let us denote by $\mathbb{U}(\boldsymbol{X})$ the set of upper-semicontinuous functions defined on $\boldsymbol{X}$. Clearly, from Proposition 7.34 of [2] and (C.1), the operator $\mathfrak{B}$ maps $\mathbb{U}(\boldsymbol{X})$ into $\mathbb{U}(\boldsymbol{X})$. Consider the sequence $(V_i)_{i\in\mathbb{N}}$ defined by $V_{i+1} = \mathfrak{B}V_i$ with $V_0 = K/\eta$. We will show that $V_i \in \mathbb{B}(\boldsymbol{X})$ for any $i \in \mathbb{N}$. By the definition of $\mathfrak{B}$ and (A.1)–(A.2), we have

$$\begin{aligned} V_1(x) &\leq \inf_{a\in\boldsymbol{A}^{\mathrm{g}}(x)}\left\{\frac{K}{K+\eta}\int_{\boldsymbol{X}} V_0(\widetilde{x})\widetilde{P}(\,\mathrm{d}\widetilde{x} \mid x, a) + \frac{1}{K+\eta}C^{\mathrm{g}}(x, a)\right\} \\ &\leq \frac{K}{K+\eta}\frac{K}{\eta} + \frac{K}{K+\eta} \\ &= \frac{K}{\eta} \\ &= V_0(x). \end{aligned}$$

From the previous inequality and since the operator $\mathfrak{B}$ is monotone, it can be easily shown by induction that the sequence $(V_i)_{i\in\mathbb{N}}$ belongs to $\mathbb{U}(\boldsymbol{X})$ and satisfies

$$V_{i+1} = \mathfrak{B}V_i \leq V_i \tag{4.9}$$

for any $i \in \mathbb{N}$. Moreover, it follows that $\sup_{x\in\boldsymbol{X}} |V_i(x)| \leq K/\eta$. Indeed, from (4.9), it follows easily that $V_i(x) \leq K/\eta$. Let us show by induction that $V_i(x) \geq -K/\eta$. Clearly, we have $V_0(x) \geq -K/\eta$. Assume that $V_i(x) \geq -K/\eta$ for $i \in \mathbb{N}$. From the definition of $\mathfrak{B}$ (see (4.6)), we have, on the one hand,

$$\inf_{a\in\boldsymbol{A}^{\mathrm{g}}(x)}\left\{\frac{K}{K+\eta}\int_{\boldsymbol{X}} V_i(\widetilde{x})\widetilde{P}(\mathrm{d}\widetilde{x} \mid x, a) + \frac{1}{K+\eta}C^{\mathrm{g}}(x, a)\right\} \geq -\frac{K}{K+\eta}\frac{K}{\eta} - \frac{K}{K+\eta} = -\frac{K}{\eta},$$

and, on the other hand,

$$\inf_{a\in\boldsymbol{A}^{\mathrm{i}}(x)}\left\{\int_{\boldsymbol{X}} V_i(\widetilde{x})Q(\mathrm{d}\widetilde{x} \mid x, a) + c^{\mathrm{i}}(x, a)\right\} \geq \inf_{a\in\boldsymbol{A}^{\mathrm{i}}(x)}\left\{\int_{\boldsymbol{X}} V_i(\widetilde{x})Q(\mathrm{d}\widetilde{x} \mid x, a)\right\} \geq -\frac{K}{\eta},$$

since c^{i} is nonnegative (recalling (A.3)). Finally, combining the two previous equations, we obtain $-K/\eta \leq V_{i+1}$. Therefore, it follows that there exists a bounded function V_∞ such that $V_i(x) \downarrow V_\infty(x)$ as $i \to \infty$ for any $x \in \boldsymbol{X}$, and so $V_\infty \in \mathbb{U}(\boldsymbol{X})$ (see [3, Theorem 4, Section 6, Chapter 4]). Now, by using (4.9), we obtain $\mathfrak{B}V_\infty \leq \mathfrak{B}V_n \leq V_n$ for any $n \in \mathbb{N}$ since the operator $\mathfrak{B}$ is monotone. This implies that $\mathfrak{B}V_\infty \leq V_\infty$. Again, from (4.9), it follows that $V_\infty \leq \mathfrak{B}V_i$ for any $i \in \mathbb{N}$. Consequently, for $x \in \mathbb{X}^{\mathrm{i}}$ and any $a^{\mathrm{g}} \in \boldsymbol{A}^{\mathrm{g}}(x)$ and $a^{\mathrm{i}} \in \boldsymbol{A}^{\mathrm{i}}(x)$, we have

$$\begin{aligned} V_\infty(x) \leq &\left\{\frac{K}{K+\eta}\int_{\boldsymbol{X}} V_i(y)\widetilde{P}(\mathrm{d}y \mid x, a^{\mathrm{g}}) + \frac{1}{K+\eta}C^{\mathrm{g}}(x, a^{\mathrm{g}})\right\} \\ &\wedge \left\{\int_{\boldsymbol{X}} V_i(y)Q(\mathrm{d}y \mid x, a^{\mathrm{i}}) + c^{\mathrm{i}}(x, a^{\mathrm{i}})\right\}. \end{aligned}$$

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

Now, by taking the limit as $i \to \infty$ in the previous equation and using the bounded convergence theorem, we obtain

$$V_\infty(x) \le \left\{ \frac{K}{K+\eta} \int_X V_\infty(y) \widetilde{P}(\mathrm{d}y \mid x, a^{\mathrm{g}}) + \frac{1}{K+\eta} C^{\mathrm{g}}(x, a^{\mathrm{g}}) \right\} \\ \wedge \left\{ \int_X V_\infty(y) Q(\mathrm{d}y \mid x, a^{\mathrm{i}}) + c^{\mathrm{i}}(x, a^{\mathrm{i}}) \right\},$$

showing that $V_\infty(x) \le \mathfrak{B} V_\infty(x)$. By using similar arguments, it is easy to show that the case in which $x \notin \mathbb{X}^{\mathrm{i}}$ leads to the same conclusion, that is, $V_\infty(x) \le \mathfrak{B} V_\infty(x)$ for all $x \in \boldsymbol{X}$. Finally, we have shown that $V_\infty = \mathfrak{B} V_\infty$ and so the function V defined by V_∞ solves (4.1).

The set $\boldsymbol{X}^{\mathrm{g}}$ coincides with

$$\left\{ x \in \boldsymbol{X} \colon V(x) = \inf_{a \in \boldsymbol{A}^{\mathrm{g}}(x)} \left\{ \frac{K}{K+\eta} \int_X V(\widetilde{x}) \widetilde{P}(\mathrm{d}\widetilde{x} \mid x, a) + \frac{1}{K+\eta} C^{\mathrm{g}}(x, a) \right\} \right\}$$

and, hence, is Borel measurable, as is $\boldsymbol{X}^{\mathrm{i}}$.

According to Proposition 7.34 of [2], for any $\varepsilon > 0$, there is a measurable map $\widetilde{\varphi}^{\mathrm{g}}$ from $\mathbf{X}$ to $\boldsymbol{A}^{\mathrm{g}}$ such that, for all $z \in \boldsymbol{X}$,

$$\widetilde{\varphi}^{\mathrm{g}}(z) \in \left\{ \widetilde{a} \in \boldsymbol{A}^{\mathrm{g}}(z) \colon \frac{K}{K+\eta} \int_X V(\widetilde{x}) \widetilde{P}(\mathrm{d}\widetilde{x} \mid x, \widetilde{a}) + \frac{1}{K+\eta} C^{\mathrm{g}}(x, \widetilde{a}) \right. \\ \left. \le \inf_{a \in \boldsymbol{A}^{\mathrm{g}}(x)} \left\{ \frac{K}{K+\eta} \int_X V(\widetilde{x}) \widetilde{P}(\mathrm{d}\widetilde{x} \mid x, a) + \frac{1}{K+\eta} C^{\mathrm{g}}(x, a) \right\} + \frac{\varepsilon}{K+\eta} \right\}.$$

The restriction of $\widetilde{\varphi}^{\mathrm{g}}$ to $\boldsymbol{X}^{\mathrm{g}}$ provides the mapping φ^{g} as required. The mapping φ^{i} is built in a similar way, working in the space $\mathbb{X}^{\mathrm{i}}$ and passing to $\boldsymbol{X}^{\mathrm{i}}$. This completes the proof.

Proposition 4.3. *Suppose that Assumptions A and (C.2) hold. Then the increasing sequence of functions* $(V_i)_{i \in \mathbb{N}}$ *defined iteratively by* $V_{i+1} = \mathfrak{B} V_i$ *with* $V_0 = -K/\eta$ *belongs to* $\mathbb{B}(\boldsymbol{X})$ *and converges to a bounded lower-semicontinuous function* V *on* $\boldsymbol{X}$ *satisfying the Bellman equation (4.1). Moreover, the corresponding sets* $\boldsymbol{X}^{\mathrm{g}}$ *and* $\boldsymbol{X}^{\mathrm{i}}$ *are measurable, and, for any* $\varepsilon \ge 0$, *there exist Borel-measurable mappings* $\varphi^{\mathrm{i}} \colon \boldsymbol{X}^{\mathrm{i}} \to \boldsymbol{A}^{\mathrm{i}}$ *and* $\varphi^{\mathrm{g}} \colon \boldsymbol{X}^{\mathrm{g}} \to \boldsymbol{A}^{\mathrm{g}}$ *respectively satisfying (4.7) and (4.8).*

Proof. According to Proposition 7.33 of [2] and by considering the sequence $(V_i)_{i \in \mathbb{N}}$ defined by $V_{i+1} = \mathfrak{B} V_i$ with $V_0 = -K/\eta$, it can be shown using the same arguments as those used in the proof of Proposition 4.2 that $|V_i(x)| \le K/\eta$, $V_{i+1} \ge V_i$, and V_i is lower semicontinuous for any $i \in \mathbb{N}$. Consequently, $(V_i)_{i \in \mathbb{N}}$ converges pointwise to a limit denoted by V_∞ which is lower semicontinuous.

Clearly, $\mathfrak{B} V_\infty \ge \mathfrak{B} V_n \ge V_n$ for any $n \in \mathbb{N}$, so $\mathfrak{B} V_\infty \ge V_\infty$. To show the reverse inequality, consider a sequence $(\overline{a}_i)_{i \in \mathbb{N}}$ of measurable mappings from $\boldsymbol{X}$ to $\boldsymbol{A}$ satisfying $\overline{a}_i(x) \in \boldsymbol{A}^{\mathrm{g}}(x) \cup \boldsymbol{A}^{\mathrm{i}}(x)$ and reaching the infimum in

$$\mathfrak{B} V_i(x) = \inf_{a \in \boldsymbol{A}^{\mathrm{g}}(x)} \left\{ \frac{K}{K+\eta} \int_X V_i(\widetilde{x}) \widetilde{P}(\mathrm{d}\widetilde{x} \mid x, a) + \frac{1}{K+\eta} C^{\mathrm{g}}(x, a) \right\} \\ \wedge \inf_{a \in \boldsymbol{A}^{\mathrm{i}}(x)} \left\{ \int_X V_i(\widetilde{x}) Q(\mathrm{d}\widetilde{x} \mid x, a) + c^{\mathrm{i}}(x, a) \right\}$$

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

for any $x \in \boldsymbol{X}$. Fix an arbitrary $x \in \boldsymbol{X}$. There exists a subsequence $(\overline{a}_{i_j}(x))_{j\in\mathbb{N}}$ of $(\overline{a}_i(x))_{i\in\mathbb{N}}$ that belongs either to $\boldsymbol{A}^{\mathrm{g}}(x)$ or $\boldsymbol{A}^{\mathrm{i}}(x)$. Consider $\overline{a}_{i_j}(x) \in \boldsymbol{A}^{\mathrm{g}}(x)$ for any $j \in \mathbb{N}$ (the other possibility can be dealt with by using the same arguments). Moreover, there is no loss of generality to assume that this subsequence converges to some $\overline{a} \in \boldsymbol{A}^{\mathrm{g}}(x)$ since $\boldsymbol{A}^{\mathrm{g}}(x)$ is compact. For $n \in \mathbb{N}$ and $j \in \mathbb{N}$ such that $n \leq i_j$, we have

$$\begin{aligned}
&\frac{K}{K+\eta}\int_{\boldsymbol{X}} V_n(\widetilde{x})\widetilde{P}(\mathrm{d}\widetilde{x} \mid x, \overline{a}_{i_j}) + \frac{1}{K+\eta}C^{\mathrm{g}}(x, \overline{a}_{i_j}) \\
&\quad \leq \frac{K}{K+\eta}\int_{\boldsymbol{X}} V_{i_j}(\widetilde{x})\widetilde{P}(\mathrm{d}\widetilde{x} \mid x, \overline{a}_{i_j}) + \frac{1}{K+\eta}C^{\mathrm{g}}(x, \overline{a}_{i_j}) \\
&\quad = \mathfrak{B}V_{i_j}(x) \\
&\quad = V_{i_j+1}(x) \\
&\quad \leq V_\infty(x),
\end{aligned} \tag{4.10}$$

where the first inequality follows from the fact that $V_{i_j} \geq V_n$. The real-valued mapping on $\boldsymbol{A}^{\mathrm{g}}(x)$ defined by $(K/(K+\eta))\int_{\boldsymbol{X}} V_n(\widetilde{x})\widetilde{P}(\mathrm{d}\widetilde{x} \mid x, \cdot) + (1/(K+\eta))C^{\mathrm{g}}(x, \cdot)$ is lower semicontinuous. Consequently, it yields $(K/(K+\eta))\int_{\boldsymbol{X}} V_n(\widetilde{x})\widetilde{P}(\mathrm{d}\widetilde{x} \mid x, \overline{a}) + (1/(K+\eta))C^{\mathrm{g}}(x, \overline{a}) \leq V_\infty(x)$ by taking the limit as j tends to ∞ in (4.10). Finally, by using the bounded convergence theorem, $(K/(K+\eta))\int_{\boldsymbol{X}} V_\infty(\widetilde{x})\widetilde{P}(\mathrm{d}\widetilde{x} \mid x, \overline{a}) + (1/(K+\eta))C^{\mathrm{g}}(x, \overline{a}) \leq V_\infty(x)$, and so $\mathfrak{B}V_\infty(x) \leq V_\infty(x)$, and, hence, $\mathfrak{B}V_\infty = V_\infty$. This shows that the lower-semicontinuous bounded function V_∞ solves (4.1). The rest of the proof is similar to that of Proposition 4.2, but is now based on Proposition 7.33 of [2] and on the fact that V is lower semicontinuous.

The next two technical lemmas are needed to construct optimal and ε-optimal control strategies in Theorem 4.1.

Lemma 4.1. *Suppose that Assumption A and either (C.1) or (C.2) hold. Let V be a bounded measurable solution of the Bellman equation (4.1). Then the following assertions hold.*

(a) *For any $x \in \boldsymbol{X}$ and $b \in \Xi$, $\int_{\boldsymbol{Y}}[C^{\mathrm{i}}(y) + V(\overline{x}(y))]\beta^b(\mathrm{d}y \mid x) \geq V(x)$.*

(b) *If, additionally, Assumption B holds then, for any $\varepsilon > 0$, there is a Markov nonrandomized policy $b^* \in \Xi$ for the controlled model $\mathcal{M}^{\mathrm{i}}$ such that, for any $x \in \boldsymbol{X}$,*

$$\int_{\boldsymbol{Y}}[C^{\mathrm{i}}(y) + V(\overline{x}(y))]\beta^{b^*}(\mathrm{d}y \mid x) \leq V(x) + \varepsilon$$

and

$$\beta^{b^*}(\{y \in Y : \overline{x}(y) \in \boldsymbol{X}^{\mathrm{g}}\} \mid x) = 1. \tag{4.11}$$

Moreover, under Assumption A, B, and (C.2), the statement of (b) can be strengthened. Indeed, it holds for $\varepsilon = 0$ for a stationary nonrandomized policy $b^* \in \Xi$.

Proof. Associated with the discrete-time MDP $\mathcal{M}^{\mathrm{i}}$, consider the cost per stage function defined on $\boldsymbol{X}_\Delta \times \boldsymbol{A}^{\mathrm{i}}_\Delta$ by $D = c^{\mathrm{i}} + \mathbf{1}_{\boldsymbol{X}\times\{\Delta\}}V$. Let $x \in \boldsymbol{X}$, and let b be an arbitrary policy for $\mathcal{M}^{\mathrm{i}}$ generating the process $(\widetilde{x}_j, \widetilde{a}_j)_{j\in\mathbb{N}}$ with initial distribution δ_x and the corresponding strategic measure $\beta^b(\cdot \mid x)$ on $(\boldsymbol{X}_\Delta \times \boldsymbol{A}^{\mathrm{i}}_\Delta)^\infty$. Let $\mathbb{E}^b_x[\cdot]$ represent the expectation with respect to this strategic measure $\beta^b(\cdot \mid x)$. Let $\tau = \inf\{j \in \mathbb{N} : \widetilde{a}_j = \Delta\}$. Then, using the bounded

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

convergence theorem and the definition of Q_Δ,

$$\lim_{m\to\infty} \mathbb{E}_x^b\left[\sum_{j=0}^{m} \mathbf{1}_{\boldsymbol{X}\times\{\Delta\}}(\widetilde{x}_j, \widetilde{a}_j)V(\widetilde{x}_j)\right] = \mathbb{E}_x^b\left[\sum_{j=0}^{\infty} \mathbf{1}_{\boldsymbol{X}\times\{\Delta\}}(\widetilde{x}_j, \widetilde{a}_j)V(\widetilde{x}_j)\right]$$
$$= \mathbb{E}_x^b[V(\widetilde{x}_\tau)\,\mathbf{1}_{\{\tau<\infty\}}]. \tag{4.12}$$

Moreover, since c^{i} is nonnegative,

$$\lim_{m\to\infty} \mathbb{E}_x^b\left[\sum_{j=0}^{m} c^{\mathrm{i}}(\widetilde{x}_j, \widetilde{a}_j)\right] = \mathbb{E}_x^b\left[\sum_{j=0}^{\infty} c^{\mathrm{i}}(\widetilde{x}_j, \widetilde{a}_j)\right], \tag{4.13}$$

by the monotone convergence theorem. Therefore, (4.12) and (4.13) yield

$$\lim_{m\to\infty} \mathbb{E}_x^b\left[\sum_{j=0}^{m} D(\widetilde{x}_j, \widetilde{a}_j)\right] = \mathbb{E}_x^b\left[\sum_{j=0}^{\infty} D(\widetilde{x}_j, \widetilde{a}_j)\right]. \tag{4.14}$$

Regarding item (a), we have $\beta^b(\{\tau < \infty\} \mid x) = 1$ since $b \in \Xi$, and so

$$\mathbb{E}_x^b\left[\sum_{j=0}^{\infty} D(\widetilde{x}_j, \widetilde{a}_j)\right] = \int_{\boldsymbol{Y}} [C^{\mathrm{i}}(y) + V(\overline{x}(y))]\beta^b(\mathrm{d}y \mid x). \tag{4.15}$$

Consider the function V_Δ defined on $\boldsymbol{X}_\Delta$ by $V_\Delta(z) = V(z)$ if $z \in \boldsymbol{X}$ and $V_\Delta(\Delta) = 0$. From (4.1), it is easy to show that V_Δ satisfies the following inequality:

$$\inf_{a\in \boldsymbol{A}^{\mathrm{i}}_\Delta(x)} \left\{D(x,a) + \int_{\boldsymbol{X}} V_\Delta(z)Q_\Delta(\mathrm{d}z \mid x,a)\right\} \ge V_\Delta(x), \qquad x \in \boldsymbol{X}_\Delta.$$

We have

$$\mathbb{E}_x^b[V_\Delta(\widetilde{x}_{m+1}) \mid \sigma\{(\widetilde{x}_j, \widetilde{a}_j)\colon j \in \mathbb{N}_m\}] = \int_{\boldsymbol{X}_\Delta} V_\Delta(z)Q_\Delta(\mathrm{d}z \mid \widetilde{x}_m, \widetilde{a}_m)$$
$$\ge V_\Delta(\widetilde{x}_m) - D(\widetilde{x}_m, \widetilde{a}_m)$$

for any $m \in \mathbb{N}$. Since V_Δ is bounded, we have

$$\mathbb{E}_x^b\left[\sum_{j=0}^{m} D(\widetilde{x}_j, \widetilde{a}_j)\right] \ge V_\Delta(x) - \mathbb{E}_x^b[V_\Delta(\widetilde{x}_{m+1})] = V(x) - \mathbb{E}_x^b[V_\Delta(\widetilde{x}_{m+1})]$$

for any $m \in \mathbb{N}$. Therefore, taking the limit as m tends to ∞ in the previous inequality and using (4.14)–(4.15), we have $\int_{\boldsymbol{Y}}[C^{\mathrm{i}}(y)+V(\overline{x}(y))]\beta^b(\mathrm{d}y \mid x) \ge V(x) - \limsup_{m\to\infty} \mathbb{E}_x^b[V_\Delta(\widetilde{x}_{m+1})]$. But, $|\mathbb{E}_x^b[V_\Delta(\widetilde{x}_m)]| \le \sup_{z\in X} |V(z)|\beta^b(\{m \le \tau\} \mid x)$, and so $\limsup_{m\to\infty} \mathbb{E}_x^b[V_\Delta(\widetilde{x}_{m+1})] = 0$ since $\beta^b(\{\tau = \infty\} \mid x) = 0$, yielding the result.

To prove item (b), we now introduce the Markov nonrandomized policy b^* for the controlled model $\mathcal{M}^{\mathrm{i}}$ defined by $b^* = (\varphi^{\mathrm{i}}_j)_{j\in\mathbb{N}}$, where, for $j \in \mathbb{N}$, φ^{i}_j is the $\boldsymbol{A}^{\mathrm{i}}_\Delta$-valued measurable mapping

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

defined on $\boldsymbol{X}_\Delta$ satisfying the following requirements.

- If $x \in \boldsymbol{X}^{\mathrm{i}}$ then

$$\varphi^{\mathrm{i}}_j(x) \in \left\{a \in A^{\mathrm{i}}(x)\colon \int_{\boldsymbol{X}} V(z)Q(\mathrm{d}z \mid x, a) + c^{\mathrm{i}}(x, a) \le V(x) + \varepsilon\left(\frac{1}{2}\right)^{j+1}\right\},$$

  which is not empty by the definition of $\boldsymbol{X}^{\mathrm{i}}$.

- If $x \in \boldsymbol{X}^{\mathrm{g}} \cup \{\Delta\}$ then $\varphi^{\mathrm{i}}_j(x) = \Delta$.

The existence of such a measurable mapping was established in Proposition 4.2 under Assumptions A and (C.1) and in Proposition 4.3 under Assumptions A and (C.2). Now, for any $j \in \mathbb{N}$ and $x \in \boldsymbol{X}_\Delta$, $D(x, \varphi^{\mathrm{i}}_j(x)) + \int_{\boldsymbol{X}_\Delta} V_\Delta(z)Q_\Delta(\mathrm{d}z \mid x, \varphi^{\mathrm{i}}_j(x)) \le V_\Delta(x) + \varepsilon(\frac{1}{2})^{j+1}$. Indeed, the previous inequality clearly holds for $x \in \boldsymbol{X}^{\mathrm{g}} \cup \{\Delta\}$, and if $x \in \boldsymbol{X}^{\mathrm{i}}$ then it follows from the definition of $\varphi^{\mathrm{i}}_j(x)$. Consequently,

$$\begin{aligned}\mathbb{E}^{b^*}_x[V_\Delta(\widetilde{x}_{m+1}) \mid \sigma\{(\widetilde{x}_j, \widetilde{a}_j)\colon j \in \mathbb{N}_m\}] &= \int_{\boldsymbol{X}_\Delta} V_\Delta(z)Q(\mathrm{d}z \mid \widetilde{x}_m, \widetilde{a}_m)\\ &\le V_\Delta(\widetilde{x}_m) - D(\widetilde{x}_m, \widetilde{a}_m) + \varepsilon\big(\tfrac{1}{2}\big)^{j+1}\end{aligned}$$

for any $m \in \mathbb{N}$. Therefore, by using the fact that V_Δ is bounded, we obtain

$$\mathbb{E}^{b^*}_x\left[\sum_{j=0}^{m} D(\widetilde{x}_j, \widetilde{a}_j)\right] \le V_\Delta(x) - \mathbb{E}^{b^*}_x[V_\Delta(\widetilde{x}_{m+1})] + \varepsilon\frac{1}{2}\,\frac{1 - (1/2)^{m+1}}{1 - 1/2}$$

and so $\limsup_{m\to\infty} \mathbb{E}^{b^*}_x[\sum_{j=0}^{m} c^{\mathrm{i}}(\widetilde{x}_j, \widetilde{a}_j)] < \infty$. Moreover, from Assumption B,

$$\{\tau = \infty\} \subset \left\{\limsup_{m\to\infty} \sum_{j=0}^{m} c^{\mathrm{i}}(\widetilde{x}_j, \widetilde{a}_j) = \infty\right\}.$$

By the monotone convergence theorem, we have $\beta^{b^*}(\{\tau < \infty\} \mid x) = 1$, implying $b^* \in \Xi$. Now, using similar arguments to those used to prove (a), we obtain

$$\int_{Y} [C^{\mathrm{i}}(y) + V(\overline{x}(y))]\beta^{b^*}(\mathrm{d}y \mid x) = \lim_{m\to\infty} \mathbb{E}^{b^*}_x\left[\sum_{j=0}^{m} D(\widetilde{x}_j, \widetilde{a}_j)\right] \le V_\Delta(x) + \varepsilon = V(x) + \varepsilon.$$

Finally, observe that $\{\tau < \infty\} \subset \{y \in \boldsymbol{Y}\colon \overline{x}(y) \in \boldsymbol{X}^{\mathrm{g}}\}$, giving the last assertion.

The proof of the last statement is similar to part (b), with a reference to Proposition 4.3.

Lemma 4.2. *Suppose that Assumption A and either (C.1) or (C.2) hold. Let $\beta \in \mathcal{P}^{\boldsymbol{Y}}$, and let V be a bounded solution of the Bellman equation (4.1). Then, for any $(x, a) \in \mathbb{K}^{\mathrm{g}}$, let*

$$\begin{aligned}&-\eta V(x) + \int_{\boldsymbol{X}}\int_{\boldsymbol{Y}} [V(\overline{x}(y)) + C^{\mathrm{i}}(y)]\beta(\mathrm{d}y \mid z)\overline{q}(\mathrm{d}z \mid x, a) - V(x)\overline{q}(\boldsymbol{X} \mid x, a) + C^{\mathrm{g}}(x, a)\\ &\quad\ge 0.\end{aligned}$$

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

Proof. From Lemma 4.1, it follows that, for any $z \in \boldsymbol{X}$, $\int_{\boldsymbol{Y}}[V(\overline{x}(y)) + C^{\mathrm{i}}(y)]\beta(\mathrm{d}y \mid z) \geq V(z)$, and so recalling that $\overline{q}$ is a positive kernel,

$$
\begin{aligned}
&-\eta V(x) + \int_{\boldsymbol{X}}\int_{\boldsymbol{Y}}[V(\overline{x}(y)) + C^{\mathrm{i}}(y)]\beta(\mathrm{d}y \mid z)\overline{q}(\mathrm{d}z \mid x, a) - V(x)\overline{q}(\boldsymbol{X} \mid x, a) + C^{\mathrm{g}}(x, a) \\
&\qquad \geq -\eta V(x) + \int_{\boldsymbol{X}} V(z)\overline{q}(\mathrm{d}z \mid x, a) - V(x)\overline{q}(\boldsymbol{X} \mid x, a) + C^{\mathrm{g}}(x, a)
\end{aligned}
$$

for any $(x, a) \in \mathbb{K}^{\mathrm{g}}$. Now the result follows from (4.1).

The next result shows the existence of optimal and ε-optimal control strategies.

Theorem 4.1. *Suppose that Assumptions A and B hold. Let V be a bounded measurable solution of the Bellman equation (4.1).*

(a) *If (C.1) or (C.2) hold then $\mathcal{V}(u, x_0) \geq V(x_0)$ for any control strategy $u = (u_n)_{n\in\mathbb{N}} \in \mathcal{U}$ with $u_n = (\psi_n, \pi_n, \gamma_n^0, \gamma_n^1)$ for $n \in \mathbb{N}^*$, and, for any $\varepsilon > 0$, there is a nonrandomized almost stationary strategy u^* such that $\mathcal{V}(u^*, x_0) \leq V(x_0) + \varepsilon$, which satisfies $\psi_n(\cdot \mid h_n) = \delta_\infty(\cdot)$, that is, the interventions occur only after the natural jumps (and maybe at the initial time moment $t = 0$).*

(b) *If (C.2) holds then there exists a nonrandomized stationary strategy u^* such that $\mathcal{V}(u^*, x_0) = V(x_0)$, which satisfies $\psi_n(\cdot \mid h_n) = \delta_\infty(\cdot)$, that is, the interventions occur only after the natural jumps (and maybe at the initial time moment $t = 0$).*

Proof. (a) If $\mathcal{V}(u, x_0) = +\infty$ then the inequality is clearly satisfied. If $\mathcal{V}(u, x_0) < +\infty$ then we have $\mathbb{P}_{x_0}^u(T_\infty < \infty) = 0$ from Proposition 4.1. Since V is bounded, Lemma 3.2 yields

$$
\begin{aligned}
\mathcal{V}(u, x_0) = &\int_{\boldsymbol{Y}} V(\overline{x}(y))u_0(\mathrm{d}y \mid x_0) + \int_{\boldsymbol{Y}} C^{\mathrm{i}}(y)u_0(\mathrm{d}y \mid x_0) \\
&+ \mathbb{E}_{x_0}^u\bigg[\int_0^{+\infty}\int_{\boldsymbol{A}^{\mathrm{g}}} \mathrm{e}^{-\eta s}\Big[-\eta V(\overline{x}(\xi_s)) + C^{\mathrm{g}}(\overline{x}(\xi_{s-}), a) \\
&\qquad + \int_{\boldsymbol{X}}\int_{\boldsymbol{Y}}\{V(\overline{x}(y)) + C^{\mathrm{i}}(y)\}\gamma^0(\mathrm{d}y \mid v, s)\overline{q}(\mathrm{d}v \mid \overline{x}(\xi_s), a) \\
&\qquad - V(\overline{x}(\xi_s))\overline{q}(\boldsymbol{X} \mid \overline{x}(\xi_s), a)\Big]\pi(\mathrm{d}a \mid s)\,\mathrm{d}s\bigg] \\
&+ \mathbb{E}_{x_0}^u\bigg[\sum_{n\in\mathbb{N}^*}\int_{(T_n, T_{n+1}]}\int_{\boldsymbol{Y}} \mathrm{e}^{-\eta s}[V(\overline{x}(y)) + C^{\mathrm{i}}(y) - V(\overline{x}(\xi_{s-}))] \\
&\qquad \times \gamma_n^1(\mathrm{d}y \mid H_n)\frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)}\bigg].
\end{aligned}
\tag{4.16}
$$

Observe that $\gamma_n^1(\mathrm{d}y \mid h_n) \in \mathcal{P}^{\boldsymbol{Y}}(\overline{x}(y_n))$ and $\gamma_n^0(\cdot \mid h_n, s, \cdot) \in \mathcal{P}^{\boldsymbol{Y}}$ for any $n \in \mathbb{N}^*$, $s \in \mathbb{R}_+^*$, and $h_n = (y_0, \theta_1, y_1, \ldots, \theta_n, y_n) \in \boldsymbol{H}_n$. Consequently, it follows from Lemma 4.1 that $\int_{\boldsymbol{Y}}[V(\overline{x}(y)) + C^{\mathrm{i}}(y) - V(\overline{x}(y_n))]\gamma_n^1(\mathrm{d}y \mid h_n) \geq 0$ for any $n \in \mathbb{N}^*$ and $h_n = (y_0, \theta_1, y_1, \ldots, \theta_n, y_n) \in \boldsymbol{H}_n$. Now, by recalling Lemma 4.2, we have

$$
\begin{aligned}
&-\eta V(\overline{x}(y_n)) + \int_{\boldsymbol{X}}\int_{\boldsymbol{Y}}[V(\overline{x}(y)) + C^{\mathrm{i}}(y)]\gamma_n^0(\mathrm{d}y \mid h_n, s, x)\overline{q}(\mathrm{d}x \mid \overline{x}(y_n), a) \\
&- V(\overline{x}(y_n))\overline{q}(\boldsymbol{X} \mid \overline{x}(y_n), a) + C^{\mathrm{g}}(\overline{x}(y_n), a) \\
&\quad \geq 0
\end{aligned}
$$

for any $n \in \mathbb{N}^*$, $s \in \mathbb{R}_+^*$, $h_n = (y_0, \theta_1, y_1, \dots, \theta_n, y_n) \in \boldsymbol{H}_n$, and $a \in \boldsymbol{A}^{\mathrm{g}}(\overline{x}(y_n))$. Observe that $\xi_{s-} = Y_n$ on the stochastic interval $]\!]T_n, T_{n+1}]\!]$ and that $\xi_{s-} = \xi_s$ on stochastic interval $]\!]T_n, T_{n+1}[\![$. Therefore, the two previous equations yield

$$\begin{aligned}
\int_0^{+\infty} \int_{\boldsymbol{A}^{\mathrm{g}}} \mathrm{e}^{-\eta s} \bigg[& -\eta V(\overline{x}(\xi_s)) + C^{\mathrm{g}}(\overline{x}(\xi_{s-}), a) \\
& + \int_{\boldsymbol{X}} \int_{\boldsymbol{Y}} \{V(\overline{x}(y)) + C^{\mathrm{i}}(y)\} \gamma^0(\mathrm{d}y \mid x, s) \overline{q}(\mathrm{d}x \mid \overline{x}(\xi_s), a) \\
& - V(\overline{x}(\xi_s)) \overline{q}(\boldsymbol{X} \mid \overline{x}(\xi_s), a) \bigg] \pi(\mathrm{d}a \mid s)\, \mathrm{d}s \\
+ \sum_{n \in \mathbb{N}^*} \int_{(T_n, T_{n+1}]} \int_{\boldsymbol{Y}} & \mathrm{e}^{-\eta s} [V(\overline{x}(y)) + C^{\mathrm{i}}(y) - V(\overline{x}(\xi_{s-}))] \\
& \times \gamma_n^1(\mathrm{d}y \mid H_n) \frac{\psi_n(\mathrm{d}s - T_n \mid H_n)}{\psi_n([s - T_n, +\infty] \mid H_n)} \\
\geq 0 \quad & \mathbb{P}_{x_0}^u\text{-almost surely,}
\end{aligned}$$

implying that $\mathcal{V}(u, x_0) \geq \int_{\boldsymbol{Y}} V(\overline{x}(y)) u_0(\mathrm{d}y \mid x_0) + \int_{\boldsymbol{Y}} C^{\mathrm{i}}(y) u_0(\mathrm{d}y \mid x_0)$. Finally, we obtain the first assertion, that is, $\mathcal{V}(u, x_0) \geq V(x_0)$, by using Lemma 4.1(a) and recalling that $u_0 \in \mathcal{P}^{\boldsymbol{Y}}(x_0)$.

Fix an arbitrary $\varepsilon > 0$, and consider the following nonrandomized almost stationary control strategy $u^* = (u_n^*)_{n \in \mathbb{N}}$ with $u_n^* = (\psi_n, \pi_n, \gamma_n^0, \gamma_n^1)$ for $n \in \mathbb{N}^*$, given by the following elements. Let $\psi_n(\cdot \mid h_n) = \delta_\infty(\cdot)$ (i.e. the interventions occur only after the natural jumps and maybe at the initial time moment $t = 0$). Set $\pi_n(\cdot \mid h_n, t) = \delta_{\varphi^{\mathrm{g}}(\overline{x}(y_n))}(\cdot)$, where, for $\widetilde{x} \in \boldsymbol{X}^{\mathrm{g}}$,

$$\varphi^{\mathrm{g}}(\widetilde{x}) \in \left\{ a \in \boldsymbol{A}^{\mathrm{g}}(\widetilde{x}) \colon \int_{\boldsymbol{X}} V(v) \overline{q}(\mathrm{d}v \mid \widetilde{x}, a) - V(\widetilde{x}) \overline{q}(\boldsymbol{X} \mid \widetilde{x}, a) + C^{\mathrm{g}}(\widetilde{x}, a) \leq \eta V(\widetilde{x}) + \frac{\eta \varepsilon}{3} \right\}$$

and, for $\widetilde{x} \in \boldsymbol{X}^{\mathrm{i}}$, $\varphi^{\mathrm{g}}(\cdot)$ is an arbitrary measurable mapping from $\boldsymbol{X}^{\mathrm{i}}$ to $\boldsymbol{A}^{\mathrm{g}}$ with $\varphi^{\mathrm{g}}(x) \in \boldsymbol{A}^{\mathrm{g}}(x)$. The existence of such a mapping follows from Proposition 4.2 under Assumptions A and (C.1) and from Proposition 4.3 under Assumptions A and (C.2). Let $\gamma_n^0(\cdot \mid h_n, \widetilde{x}) = \beta^{b^*}(\cdot \mid \widetilde{x})$, with the policy $b^* \in \Xi$ introduced in Lemma 4.1(b) and satisfying the inequality

$$\int_{\mathbf{Y}} [C^{\mathrm{i}}(y) + V(\overline{x}(y))] \beta^{b^*}(\mathrm{d}y \mid x) \leq V(x) + \min\left\{1, \frac{\eta}{K}\right\} \frac{\varepsilon}{3} \tag{4.17}$$

and (4.11) for any $x \in X$. Consider γ_n^1 as an arbitrary stochastic kernel on $\boldsymbol{Y}$ given $\boldsymbol{H}_n$ satisfying $\gamma_n^1(\cdot \mid h_n) \in \mathcal{P}^{\boldsymbol{Y}^*}(\overline{x}(y_n))$ for any $h_n = (y_0, \theta_1, \dots, \theta_n, y_n) \in \boldsymbol{H}_n$ with $\overline{x}(y_n) \in \mathbb{X}^{\mathrm{i}}$. Finally, set $u_0 = \beta^{b^*}(\cdot \mid x_0)$.

Firstly, it follows that $\mathbb{P}_{x_0}^{u^*}(\Theta_{n+1} \in \Gamma_\theta \mid \mathcal{F}_{T_n}) = \int_{\Gamma_\theta} \lambda_n(\boldsymbol{X}, H_n, t) \mathrm{e}^{-\Lambda_n(\boldsymbol{X}, H_n, t)}\, \mathrm{d}t$, from Remark 2.1 and the definition of G_n (see (2.1)). However, (A.1) ensures that $\lambda_n(\boldsymbol{X}, H_n, t)$ is uniformly bounded by the constant K and so $\mathbb{P}_{x_0}^{u^*}(T_\infty < \infty) = 0$. Consequently, according to

(4.16) and the definition of the strategy u^*, we have

$$\begin{aligned}
&\mathcal{V}(u^*, x_0)\\
&\quad = \int_{\boldsymbol{Y}} V(\overline{x}(y))\beta^{b^*}(\mathrm{d}y \mid x_0) + \int_{\boldsymbol{Y}} C^{\mathrm{i}}(y)\beta^{b^*}(\mathrm{d}y \mid x_0)\\
&\qquad + \mathbb{E}^{u^*}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s}[-\eta V(\overline{x}(\xi_s)) + C^{\mathrm{g}}(\overline{x}(\xi_{s-}), \varphi^{\mathrm{g}}(\overline{x}(\xi_{s-})))]\mathrm{d}s\bigg]\\
&\qquad + \mathbb{E}^{u^*}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s}\bigg[\int_{\boldsymbol{X}}\int_{\boldsymbol{Y}}\{V(\overline{x}(y)) + C^{\mathrm{i}}(y)\}\beta^{b^*}(\mathrm{d}y \mid x)\overline{q}(\mathrm{d}x \mid \overline{x}(\xi_s), \varphi^{\mathrm{g}}(\overline{x}(\xi_{s-})))\\
&\qquad\qquad - V(\overline{x}(\xi_s))\overline{q}(\boldsymbol{X} \mid \overline{x}(\xi_s), \varphi^{\mathrm{g}}(\overline{x}(\xi_{s-})))\bigg]\mathrm{d}s\bigg].
\end{aligned}$$

Now, from inequality (4.17) we have

$$\begin{aligned}
\mathcal{V}(u^*, x_0) \le V(x_0) + \frac{\varepsilon}{3} + \frac{\eta\varepsilon}{3K}\mathbb{E}^{u}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s}\overline{q}(\boldsymbol{X} \mid \overline{x}(\xi_s), \varphi^{\mathrm{g}}(\overline{x}(\xi_{s-})))\,\mathrm{d}s\bigg]&\\
+ \mathbb{E}^{u}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s}\mathcal{W}(s)\,\mathrm{d}s\bigg],& \qquad (4.18)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{W}(s) &= -\eta V(\overline{x}(\xi_s)) + C^{\mathrm{g}}(\overline{x}(\xi_{s-}), \varphi^{\mathrm{g}}(\overline{x}(\xi_s)))\\
&\quad + \int_{\boldsymbol{X}} V(x)\overline{q}(\,\mathrm{d}x \mid \overline{x}(\xi_s), \varphi^{\mathrm{g}}(\overline{x}(\xi_s))) - V(\overline{x}(\xi_s))\overline{q}(\boldsymbol{X} \mid \overline{x}(\xi_s), \varphi^{\mathrm{g}}(\overline{x}(\xi_s))). \qquad (4.19)
\end{aligned}$$

Observe that, since V is bounded and by Assumption A, it follows that $|\mathcal{W}|$ is uniformly bounded by a constant $K_{\mathcal{W}}$. By Fubini's theorem,

$$\begin{aligned}
\mathbb{E}^{u}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s}\mathcal{W}(s)\,\mathrm{d}s\bigg] \le \sum_{n\in\mathbb{N}^*}\mathbb{E}^{u}_{x_0}\bigg[\mathbf{1}_{\boldsymbol{X}^{\mathrm{g}}}(\overline{x}(Y_n))\int_{[T_n, T_{n+1})} \mathrm{e}^{-\eta s}\mathcal{W}(s)\,\mathrm{d}s\bigg]&\\
+ K_{\mathcal{W}}\sum_{n\in\mathbb{N}^*}\mathbb{P}^{u}_{x_0}(\overline{x}(Y_n) \in \boldsymbol{X}^{\mathrm{i}}).& \qquad (4.20)
\end{aligned}$$

Now, observe that $\mathbb{P}^{u}_{x_0}(\overline{x}(Y_{n+1}) \in \boldsymbol{X}^{\mathrm{i}} \mid \mathcal{F}_{T_n}) = G_n(\Gamma^{\mathrm{i}}_y \times \mathbb{R}^*_+ \mid H_n)$, where $\Gamma^{\mathrm{i}}_y = \{y \in \boldsymbol{Y} : \overline{x}(y) \in \boldsymbol{X}^{\mathrm{i}}\}$. However, $\gamma^0_n(\Gamma^{\mathrm{i}}_y \mid H_n, t, x) = \beta^{b^*}(\Gamma^{\mathrm{i}}_y \mid x) = 0$ for any $x \in \boldsymbol{X}$ according to (4.11), and so from the definition of G_n we have $\mathbb{P}^{u}_{x_0}(\overline{x}(Y_{n+1}) \in \boldsymbol{X}^{\mathrm{i}} \mid \mathcal{F}_{T_n}) = 0$, implying that $\mathbb{P}^{u}_{x_0}(\overline{x}(Y_{n+1}) \in \boldsymbol{X}^{\mathrm{i}}) = 0$ for any $n \in \mathbb{N}^*$. Moreover, $\mathbb{P}^{u}_{x_0}(\overline{x}(Y_1) \in \boldsymbol{X}^{\mathrm{i}}) = \beta^{b^*}(\Gamma^{\mathrm{i}}_y \mid x_0)$ according to (4.11). Consequently, from (4.20), it follows that

$$\mathbb{E}^{u}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s}\mathcal{W}(s)\,\mathrm{d}s\bigg] \le \sum_{n\in\mathbb{N}^*}\mathbb{E}^{u}_{x_0}\bigg[\mathbf{1}_{\boldsymbol{X}^{\mathrm{g}}}(\overline{x}(Y_n))\int_{[T_n, T_{n+1})} \mathrm{e}^{-\eta s}\mathcal{W}(s)\,\mathrm{d}s\bigg],$$

and so, from (4.19) and the definition of φ^{g} on $\boldsymbol{X}^{\mathrm{g}}$, it follows that

$$\mathbb{E}^{u}_{x_0}\bigg[\int_0^{+\infty} \mathrm{e}^{-\eta s}\mathcal{W}(s)\,\mathrm{d}s\bigg] \le \frac{\varepsilon}{3}. \qquad (4.21)$$

Combining (4.18), (4.21), and (A.1), we obtain $\mathcal{V}(u^*, x_0) \le V(x_0) + \varepsilon$.

(b) The proof is similar to that of (a), now using Proposition 4.3 and the last statement of Lemma 4.1.

https://doi.org/10.1239/aap/1427814583 Published online by Cambridge University Press

In the next result we will show the existence of uniformly optimal and ε-optimal control strategies.

Corollary 4.1. *The following assertions hold.*

(a) *If Assumptions A, B, and (C.1) hold then, for any $\varepsilon > 0$, there exists a nonrandomized, almost stationary, uniformly ε-optimal strategy u^* with $\psi_n(\cdot \mid h_n) = \delta_\infty(\cdot)$, that is, the interventions occur only after the natural jumps (and maybe at the initial time moment $t = 0$).*

(b) *If Assumptions A, B, and (C.2) hold then there is a nonrandomized stationary uniformly optimal strategy u^* satisfying $\psi_n(\cdot \mid h_n) = \delta_\infty(\cdot)$, that is, the interventions occur only after the natural jumps (and maybe at the initial time moment $t = 0$).*

(c) *In either case, $\inf_{u \in \mathcal{U}} \mathcal{V}(u, x_0) = V(x_0)$, and the Bellman equation (4.1) has a unique bounded measurable solution.*

Proof. From the proof of Theorem 4.1, the control strategies u^* do not depend on the initial condition x_0.

Roughly speaking, one should apply the gradual action $\varphi^{\mathrm{g}}(x)$ if the current state is $x \in \boldsymbol{X}^{\mathrm{g}}$, and one should apply the impulsive action $\varphi^{\mathrm{i}}(x)$ if $x \in \boldsymbol{X}^{\mathrm{i}}$ (see Proposition 4.2).

The impulsive actions serve to push the original process away from the set $\boldsymbol{X}^{\mathrm{i}}$ which, e.g. contains the states with very large cost rates $C^{\mathrm{g}}(x, a)$. It is intuitively obvious that these actions must be applied as quickly as possible, namely, immediately after any natural jump leading to a new state $x \in \mathbf{X}^{\mathrm{i}}$. If $x \notin \boldsymbol{X}^{\mathrm{i}}$ then there is no need to intervene, which explains the choice of $\psi_n(\cdot \mid h_n) = \delta_\infty(\cdot)$. The initial state x_0 of the original process can also be in $\boldsymbol{X}^{\mathrm{i}}$. In this case, the controller needs to apply an initial sequence of impulses corresponding to the component u_0 of a strategy. Therefore, it is inevitable to have $\theta_1 = 0$: $\xi_{0-}(\omega) = y_0$ corresponds to the given initial state x_0 before any action is applied, and $\xi_0(\omega) = y_1$ results in an initial intervention to push the original process away from the set $\boldsymbol{X}^{\mathrm{i}}$

References

[1] Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

[2] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case* (Math. Sci. Eng. **139**). Academic Press, New York.

[3] Bourbaki, N. (1971). *Éléments de Mathématique. Topologie Générale. Chapitres 1 à 4*. Hermann, Paris.

[4] Brémaud, P. (1981). *Point Processes and Queues*. Springer, New York.

[5] Davis, M. H. A. (1993). *Markov Models and Optimization* (Monogr. Statist. Appl. Prob. **49**). Chapman & Hall, London.

[6] De Leve, G. (1964). *Generalized Markovian Decision Processes. Part I: Model and Method* (Math. Centre Tracts **3**). Mathematisch Centrum, Amsterdam.

[7] De Leve, G. (1964). *Generalized Markovian Decision Processes. Part II: Probabilistic Background* (Math. Centre Tracts **4**). Mathematisch Centrum, Amsterdam.

[8] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes. Theory and Applications* (Stoch. Modelling Appl. Prob. **62**). Springer, Berlin.

[9] Guo, X., Hernández-Lerma, O. and Prieto-Rumeau, T. (2006). A survey of recent results on continuous-time Markov decision processes. With comments and a rejoinder by the authors. *Top* **14,** 177–261.

[10] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes. Basic Optimality Criteria* (Appl. Math. (New York) **30**). Springer, New York.

[11] Hordijk, A. and van der Duyn Schouten, F. A. (1983). Average optimal policies in Markov decision drift processes with applications to a queueing and a replacement model. *Adv. Appl. Prob.* **15,** 274–303.

[12] Hordijk, A. and van der Duyn Schouten, F. A. (1984). Discretization and weak convergence in Markov decision drift processes. *Math. Operat. Res.* **9,** 112–141.

[13] Hordijk, A. and van der Duyn Schouten, F. (1985). Markov decision drift processes: conditions for optimality obtained by discretization. *Math. Operat. Res.* **10,** 160–173.
[14] Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. The Technology Press of M.I.T., Cambridge, MA.
[15] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon–Nikodým derivatives, representation of martingales. *Z. Wahrscheinlichkeitsth.* **31,** 235–253.
[16] Jacod, J. (1979). *Calcul Stochastique et Problèmes de Martingales* (Lecture Notes Math. **714**). Springer, Berlin.
[17] Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line. The Dynamic Approach*. Springer, New York.
[18] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games* (ICP Adv. Texts Math. **5**). Imperial College Press, London.
[19] Van der Duyn Schouten, F. A. (1983). *Markov Decision Processes With Continuous Time Parameter* (Math. Centre Tracts **164**). Mathematisch Centrum, Amsterdam.
[20] Yushkevich, A. A. (1983). Continuous time Markov decision processes with interventions. *Stochastics* **9,** 235–274.
[21] Yushkevich, A. A. (1986). Markov decision processes with both continuous and impulsive control. In *Stochastic Optimization* (Kiev, 1984; Lecture Notes Control Inf. Sci. **81**), Springer, Berlin, pp. 234–246.
[22] Yushkevich, A. A. (1987). Bellman inequalities in Markov decision deterministic drift processes. *Stochastics* **23,** 25–77.
[23] Yushkevich, A. A. (1989). Verification theorems for Markov decision processes with controllable deterministic drift and gradual and impulsive controls. *Theory Prob. Appl.* **34,** 474–496.