DIGITALCOMMONS
@WAYNESTATE

Wayne State University

Mathematics Faculty Research Publications

Mathematics

5-30-2018

Discrete-Time Hybrid Control in Borel Spaces

Héctor Jasso-Fuentes
CINVESTAV-IPN, hjasso@math.cinvestav.mx

José-Luis Menaldi
Wayne State University, menaldi@wayne.edu

Tomás Prieto-Rumeau
National University of Distance Education, tprieto@ccia.uned.es

Follow this and additional works at: https://digitalcommons.wayne.edu/mathfrp

Part of the Control Theory Commons, and the Dynamic Systems Commons

Recommended Citation

Jasso-Fuentes, H., Menaldi, JL. & Prieto-Rumeau, T.: Discrete-Time Hybrid Control in Borel Spaces. Appl Math Optim (2018). https://doi.org/10.1007/s00245-018-9503-z

This Article is brought to you for free and open access by the Mathematics at DigitalCommons@WayneState. It has been accepted for inclusion in Mathematics Faculty Research Publications by an authorized administrator of DigitalCommons@WayneState.

Published in: Appl Math Optim, 2018, pp. 1-33. https://doi.org/10.1007/s00245-018-9503-z

Discrete-Time Hybrid Control in Borel Spaces*

Héctor Jasso-Fuentes[1] José-Luis Menaldi[2] Tomás Prieto-Rumeau[3]

30 May 2018

Abstract

A discrete-time hybrid control model with Borel state and action spaces is introduced. In this type of models, the dynamic of the system is composed by two sub-dynamics affecting the evolution of the state; one is of a standard-type that runs almost every time and another is of a special-type that is active under special circumstances. The controller is able to use two different type of actions, each of them is applied to each of the two sub-dynamics, and the activations of these sub-dynamics are possible according to an *activation rule* that can be handled by the controller. The aim for the controller is to find a control policy, containing a mix of actions (of either standard- or special-type), with the purpose of minimizing an infinite-horizon discounted cost criterion whose discount factor is dependent on the state-action history and may be equal one at some stages. Two different sets of conditions are proposed to guarantee (i) the finiteness of the cost criterion, (ii) the characterization of the optimal value function and (iii) the existence of optimal control policies; to do so, we employ the dynamic programming approach. A useful characterization that signalizes the accurate times between changes of sub-dynamics in terms of the so-named contact set is also provided. Finally, we introduce two examples that illustrate our results and also show that control models such as discrete-time impulse control models and discrete-time switching control models become special cases of our present hybrid model.

2010 Mathematics Subject Classification: 93E20, 34A38, 60J05
Keywords and phrases: *Hybrid control systems, Markov decision processes, dynamic programming.*

1 Introduction

The way in which a control may affect the evolution of a dynamic system can be very complex, particularly when some digital and analogue components interact together. In the past decades, the concept of a *hybrid system* has been used to handle the situation where these components (digital and analogue) play important roles in the problem under study; in fact, in the literature, the same token "hybrid system" has been used to represent a wide variety of distinct cases, covering almost all possible real situations —see, for instance, Branicky [12], Goebel et. al. [10], Lygeros [18], Yin & Zhu [30].

The idea behind a hybrid system is a so-called event-driven evolution, i.e., under normal circumstances a *standard*-type sub-dynamic is a good description of the real phenomenon, but some events may occur (due to internal or exogenous causes) and the model becomes invalid, which forces the "modeler" to reconsider the data of the problem; consequently, the dynamic may undergo structural modifications, i.e., from time

*This research was funded in part by CONACyT fellowship 266252 "Estancias sabáticas en el extranjero 2015'', by CONACyT grant 238045, and by grant MTM2016-75497-P from the Spanish Ministerio de Economía y Competitividad.
[1]Departmento de Matemáticas, CINVESTAV-IPN, México D.F, 07000, México. hjasso@math.cinvestav.mx
[2]Department of Mathematics, Wayne State University, Detroit, MI, 48202, USA menaldi@wayne.edu)
[3]Department of Statistics and Operations Research. UNED, Madrid, Spain. tprieto@ccia.uned.es

to time the law of evolution may suffer deep changes. More specifically, in a hybrid model a standard sub-dynamic is running under almost any situations, but under some extreme circumstances, a *special*-type sub-dynamic becomes active, overruling the standard evolution. A change of sub-dynamic may produce a structural modification in the system and, at the same time, an opportunity for an instantaneous and sizeable change in the state of the system.

The way the control is applied to these models is a little subtle. Namely, in a general setting, two usual-type of controls are considered, each one (and only one) is applied to one sub-dynamic, but there is another control that can lead the activation of the sub-dynamics; that is, it determines which of them is active. Such control can be regarded as of an unusual type in the sense that its range is actually the activation of the sub-dynamics. This control is not always allowed to be triggered arbitrarily; in fact, its activation depends on the location of state of the system.

Another way to explain the activation of the sub-dynamics is from the point of view of an *activation rule*. This rule determines which sub-dynamic must be active based on the location of the state of the system and also based on the interest of the controller —when it is possible to do it. Indeed, the rule basically obligates the activation of the standard or the special sub-dynamic when the state is located at some subset of the state space, but in some other regions, there is a flexibility to change such sub-dynamics in accordance with the controller's selection.

On the other hand, the state of the system not only stands for a "usual" description of the phenomenon under study, but also has a record keeping mechanism. Specifically, the state is represented as a pair, where the first component describes the standard evolution of the system (fast-type variable) and the second one acts as a variable that records the structural changes (slow-type variable). The description of the state of the system as well as the corresponding control will be discussed with more details in later sections.

Summing-up, the system is composed by: (1) a state variable with two components (fast-type or slow-type); (2) two sub-dynamics, initially independent of each other, that are tied together by an activation rule (or unusual control) highly dependent on the state position, and (3) a control with two components, one component applied for each sub-dynamic. With this description, the aim for the controller is to find a control policy, containing a mix of actions (of either standard- or special-type), with the purpose of minimizing an infinite-horizon discounted cost criterion with discount factor dependent on the state-action history and that may equal one at some stages.

Hybrid control systems have been studied in continuous-time models —see, for instance, the works of Bensoussan & Menaldi [7, 8], Borkar et al. [11], and Branicky et. al. [13]. To the best of our knowledge, the discrete-time case has been studied in Abate et al. [1, 2] and Summers & Lygeros [28], under the special case of reachability/avoiding control defined in Euclidian spaces. Recently, these problems have been studied in Borel spaces with average cost criteria in Jasso-Fuentes, et. al [16]. Finally, there exist several references focused on special cases of hybrid models in continuous- and discrete-time, e.g., impulsive control problems —see Bensoussan [4], Bensoussan & Lions [5, 6], Menaldi [19], Robin [24, 25], Stettner [26, 27], switching control problems —see Bensoussan and Lions [5], Menaldi & Blankenship [20], Zhang et al. [31] , and standard control problems —see Bensoussan [4], Hernández-Lerma & Lasserre [14, 15], Puterman [21], and the references therein.

As it is shown in Section 5, hybrid control models include discrete-time "standard" optimal control models of the type given in Bensoussan [4], Bertsekas & Shreve [9], Hernández-Lerma & Lasserre [14, 15], and Puterman [21]. Furthermore, even when the hypotheses are similar to the standard models, the way to analyze optimality is quite different because of the inclusion of the unusual control. It is also important to say that the nature of these classes of models leads us to work with a state-action dependent discount factor with values possibly equal one at some stages.

Under our perspective, this paper has further novelties: (1) Our setting includes almost all possible ways of control (regular control, impulsive and switching-type controls) in a single model, which is defined in *general* state-action spaces. (2) Our criterion is set on an infinite horizon (but other cases can be accommodated) whose discount factor is non-constant and may depend on the current state and actions. Perhaps the most remarkable distinction of our model is the fact that the cost associated with transitions due to the special sub-dynamic are assumed to be occur instantaneously (in time), without any discount (i.e., with discount factor one); and this situation has not been seen (by us) in other papers. Certainly, this new possibility produces discontinuities in our model that require to be addressed. Our main result can be summarized as follows: under suitable assumptions the optimal cost is the unique solution to the dynamic programming equation and there exists an optimal feedback policy (Theorems 3.7 and 3.13), but there are several aspects, consequences and details that are expressed in several Propositions within the text, as required in a mathematical/theoretic paper.

This paper is divided in five sections: In the next section we introduce the dynamics of the model, the different type of control policies we are dealing with, and the discounted-type optimality criterion to be optimized. We also give sufficient and necessary conditions that ensure finiteness of this criterion. In Section 3 we give solution to our control problem through the existence of optimal control policies under two different sets of hypotheses. Furthermore, we give a characterization of the optimal value function viewed as the solution of a certain functional equation (the dynamic programming equation). The last part of the section contains a useful characterization of the unusual control that signalizes the optimal region to apply a change of sub-dynamics. In Section 4 we provide two useful applications: one is about a consumption-investment problem with market modes and the other is related to a manufacturing-production problem. As we shall see, the use of "conventional" control models (e.g. impulsive, switching or standard control models) is not sufficient to give solution to these problems, but by using hybrid control models this solution is possible. We conclude this work with Section 5, in which well-known control models such as impulsive, switching, and standard control can be regarded as special cases of our hybrid control model.

Notation and terminology Throughout this paper:

- Any metric space Z will be endowed with its Borel σ-algebra $\mathcal{B}(Z)$ and measurability (of sets and functions) will be always referred to the corresponding Borel σ-algebras.
- Given some metric space Z, the family of nonnegative measurable functions $u : Z \to [0, \infty)$ will be denoted by $\mathbb{M}^+(Z)$, while the family of nonnegative bounded measurable functions $u : Z \to [0, \infty)$ (hence, with $||u|| = \sup_{x \in Z} |u(z)| < \infty$) will be denoted by $\mathbb{B}^+(Z)$.
- A function $f : Z \to (-\infty, +\infty]$ is said to be lower semicontinuous when $\liminf_{y \to z} f(y) \geq f(z)$ for all $z \in Z$. The family of nonnegative lower semicontinuous functions $f : Z \to [0, \infty)$ is denoted by $\mathbb{L}^+(Z)$.
- We recall that a Borel space is a measurable subset of a complete and separable metric space.
- We make the convention that a product of real numbers $\prod_{i \in S} x_i$ over an empty set S equals one, while a sum $\sum_{i \in S} x_i$ over an empty set S equals zero.
- The notation $\delta_x(\cdot)$ and $\mathbf{1}_C(\cdot)$ will mean the Dirac measure concentrated on the point $x \in Z$ and the indicator function of a set $C \in \mathcal{B}(Z)$, respectively.
- For a given set D, we denote by $\bar{D}$ its closure and by $\mathring{D}$ its interior.

2 Model definition

The state and action spaces The *state space* X of a discrete-time hybrid system is the product $X = X^f \times X^s$ of two Borel spaces, where the components $x^f \in X^f$ and $x^s \in X^s$ are called the fast[4] and slow[5] states, respectively. The *action space* A is a Borel space and it is the union of two disjoint measurable subsets: $A = V^f \cup V^s$. The sets V^f and V^s are referred to as the fast and the slow action sets, respectively.

State-action pairs The set of feasible state-action pairs is given by a measurable set $\mathbb{K} \subseteq X \times A$ with nonempty X-sections, which are denoted by $(x^f, x^s) \mapsto A(x^f, x^s) \subseteq A$ for each $(x^f, x^s) \in X$. We assume further the existence of two measurable sets

$$D^\wedge \subseteq D^\vee \subseteq X$$

such that

(a) $A(x^f, x^s) \cap V^f = \emptyset$ when $(x^f, x^s) \in D^\wedge$, meaning that when the state of the system is in $D^\wedge$, the controller must necessarily choose an action in V^s (a slow action); and

(b) $A(x^f, x^s) \cap V^s = \emptyset$ when $(x^f, x^s) \in X \setminus D^\vee$, meaning that when the state of the system is outside $D^\vee$, the controller must necessarily choose an action in V^f (a fast action).

We assume that $\mathbb{K}$ contains the graph of some measurable function from X to A. Hence, the family $\mathbb{F}$ of measurable functions $\mathtt{f} : X \to A$ such that $\mathtt{f}(x^f, x^s) \in A(x^f, x^s)$ for all $(x^f, x^s) \in X$ is nonempty.

Thus, the sets $D^\wedge$ and $D^\vee$ will partially determine the activation rule of the sub-dynamics, since inside $D^\wedge$ and outside $D^\vee$ the controller is *forced* to choose an action of a specific nature (fast or slow). In contrast, when the state of the system (x^f, x^s) is in $D^\vee \setminus D^\wedge$, the controller will not have an a priori restriction on the nature of his actions.

Dynamic of the system The dynamic is composed by two sub-dynamics: one sub-dynamic is of a standard type[6], and it only affects the fast states $x^f \in X^f$ through the stochastic transition kernel

$$Q^f : \mathcal{B}(X^f) \times \big(\mathbb{K} \cap (X \times V^f)\big) \mapsto [0, 1],$$

while the other sub-dynamic is of a special type[7] and it produces a transition of both components $(x^f, x^s) \in X$ following the stochastic kernel

$$Q^s : \mathcal{B}(X) \times \big(\mathbb{K} \cap (X \times V^s)\big) \mapsto [0, 1].$$

Summarizing, the whole dynamic is given by

$$\mathbf{Q}(dy^f \times dy^s | x^f, x^s, a) = \begin{cases} Q^f(dy^f | x^f, x^s, a)\delta_{x^s}(dy^s) & \text{if } a \in V^f, \\ Q^s(dy^f \times dy^s | x^f, x^s, a) & \text{if } a \in V^s. \end{cases} \tag{2.1}$$

[4]a.k.a. continuous or regular
[5]a.k.a. discrete or impulsive
[6]so-called usual or traditional sub-dynamic
[7]so-called impulse-type or event-driven sub-dynamic

Control policies Define $H_0 = X$ and $H_k = \mathbb{K}^k \times X$ for $k \geq 1$, and let $H_\infty = \mathbb{K}^\infty$, all endowed with the corresponding product σ-algebras. The history up to step k is

$$h_k = \big(x_0^f, x_0^s, a_0, \cdots, x_{k-1}^f, x_{k-1}^s, a_{k-1}, x_k^f, x_k^s\big) \in H_k.$$

A control policy is a sequence $\{\nu_k\}_{k\geq 0}$ of transition probability measures on A given H_k such that $\nu_k\big(A(x_k^f, x_k^s)|h_k\big) = 1$ for all $h_k \in H_k$. In particular, we necessarily have

$$\nu_k\big(A(x_k^f, x_k^s) \cap V^s|h_k\big) = 1 \quad \text{if } (x_k^f, x_k^s) \in D^\wedge, \text{ and}$$

$$\nu_k\big(A(x_k^f, x_k^s) \cap V^f|h_k\big) = 1 \quad \text{if } (x_k^f, x_k^s) \in X \setminus D^\vee.$$

We denote by Π the set of admissible control policies.

By the Ionescu-Tulcea theorem, for any initial state $x = (x^f, x^s) \in X$ and any policy $\nu \in \Pi$ there exists a unique probability measure on H_∞, denoted by P_x^ν, which models the controlled dynamic system under ν. Its expectation operator is denoted by E_x^ν.

If there is some $\mathtt{f} \in \mathbb{F}$ such that the policy $\nu \in \Pi$ satisfies $\nu_k(\cdot|h_k) = \delta_{\mathtt{f}(x_k^f, x_k^s)}(\cdot)$ for any $h_k \in H_k$ and $k \geq 0$, then we say that ν is a *deterministic stationary* policy. In what follows, we will identify the set of such policies with $\mathbb{F}$. Hence, we have $\mathbb{F} \subseteq \Pi$.

Remark 2.1. **(a)** *The dynamic system can be also formulated in an equivalent way by means of two measurable functions* $F : X \times V^f \times S \to X^f$ *and* $G : X \times V^s \times S \to X$, *with* S *a Borel space, where*

$$\underbrace{\big(x_{k+1}^f, x_{k+1}^s\big) = \big(F(x_k^f, x_k^s, a_k, w_k), x_k^s\big)}_{\textit{standard sub-dynamic}} \quad \textit{if} \;\; a_k \in V^f, \tag{2.2}$$

or

$$\underbrace{\big(x_{k+1}^f, x_{k+1}^s\big) = G\big(x_k^f, x_k^s, a_k, w_k\big)}_{\textit{special sub-dynamic}} \quad \textit{if} \;\; a_k \in V^s, \tag{2.3}$$

and where $\{w_k\}$ *is a sequence of i.i.d. random variables on* S*; see Proposition 8.6 in Kallenberg [17].*

(b) *Intuitively, our hybrid dynamic model may be regarded as a two time-scales model, in which the* fast *sub-dynamic has an evolution according to* Q^f *in* (2.1) *—or* (2.2)*— whereas the* slow *sub-dynamic is driven by* Q^s *in* (2.1) *—or* (2.3)*—. We warn the reader that our model differs from some other models also named two time-scales (see for instance Yin & Zhang [29]), in which the attributes fast and slow are based on a small parameter* $\varepsilon > 0$*. The study of the latter models is mainly based on the singular perturbation theory.*

Time component We will also consider a sequence $\{\mathtt{t}_k : k \geq 0\}$ of measurable functions on $\mathbb{K}^\infty$ taking values in $\mathbb{N}$, that will represent the number of times that, previous to k, an action in V^f has been taken. At this point, we will use the notation

$$\omega = \big(x_0^f, x_0^s, a_0, \cdots, x_k^f, x_k^s, a_k, \cdots\big)$$

for an element of $H_\infty = \mathbb{K}^\infty$. Given arbitrary $\omega \in H_\infty$, we put $\mathbf{t}_0(\omega) = 0$ and, for each $k \geq 1$, we let

$$\mathtt{t}_k(\omega) = \sum_{j=0}^{k-1} \mathbf{1}_{V^f}(a_j). \tag{2.4}$$

We assume that when the standard sub-dynamic is used (that is, an action in V^f is taken) then the "natural time" component increases by one; in other words, a time unit passes. On the contrary, when the special sub-dynamic is used (that is, an action in V^s is taken) then the "natural time" does not change, and this is interpreted as an *instantaneous* transition. In this manner, $\mathtt{t}_k$ will represent the "natural time" when the system is in x_k, after k transitions.

Definition 2.2. *A sample path $\omega \in H_\infty$ such that $\lim_{k\to\infty} \mathtt{t}_k(\omega) < \infty$ will be called* explosive.

The above definition is coherent with the corresponding continuous-time terminology. Hence, a sample path $\omega \in H_\infty$ is explosive if and only if there exists some k_0 such that $a_k \in V^s$ for all $k \geq k_0$. Equivalently, a sample path $\omega \in H_\infty$ is non-explosive if and only if $a_k \in V^f$ for infinitely many $k \geq k_0$.

Optimality criterion We will consider a discounted cost optimality criterion with varying discount factor. More precisely, we will consider a running cost function $\mathbf{c} : \mathbb{K} \to [0,\infty)$ which will be written

$$\mathbf{c}(x^f, x^s, a) = c(x^f, x^s, a)\mathbf{1}_{V^f}(a) + \ell(x^f, x^s, a)\mathbf{1}_{V^s}(a),$$

with

$$c : \mathbb{K} \cap (X \times V^f) \mapsto [0,\infty) \quad \text{and} \quad \ell : \mathbb{K} \cap (X \times V^s) \mapsto [0,\infty)$$

interpreted as the running cost functions for the standard and the special sub dynamics, respectively. The discount factor function is $\alpha : \mathbb{K} \to [0,1]$. We will assume that both $\mathbf{c}$ and α are measurable.

Given an initial state $(x^f, x^s) \in X$ and a control policy $\nu \in \Pi$ we define

$$J(x^f, x^s, \nu) = E^\nu_{x^f,x^s}\Big[\sum_{k=0}^{\infty} \mathbf{c}(x^f_k, x^s_k, a_k) \prod_{i=0}^{k-1} \alpha(x^f_i, x^s_i, a_i)\Big]. \tag{2.5}$$

The optimal discounted cost function is then defined as

$$J^*(x^f, x^s) = \inf_{\nu\in\Pi} J(x^f, x^s, \nu) \quad \text{for } (x^f, x^s) \in X, \tag{2.6}$$

and we will say that a policy $\nu^* \in \Pi$ is optimal when

$$J(x^f, x^s, \nu^*) = J^*(x^f, x^s) \quad \text{for all } (x^f, x^s) \in X. \tag{2.7}$$

Observe that the discounted cost function $J(\cdot, \nu)$ is well defined but it might be infinite. Furthermore, as a consequence of the definition of $J(x^f, x^s, \nu)$, it follows that the discount factor applied at step $k \geq 1$ depends on the previous history of the process $(h_{k-1}, a_{k-1}) = (x^f_0, x^s_0, a_0, \ldots, x^f_{k-1}, x^s_{k-1}, a_{k-1})$.

Remark 2.3. **(a)** *We can incorporate a "current" discount factor $\alpha(x^f_k, x^s_k, a_k)$ at time k by simply considering the cost function $\overline{\mathbf{c}} = \mathbf{c}\alpha$, with*

$$E^\nu_{x^f,x^s}\Big[\sum_{k=0}^{\infty} \mathbf{c}(x^f_k, x^s_k, a_k) \prod_{i=0}^{k} \alpha(x^f_i, x^s_i, a_i)\Big] = E^\nu_{x^f,x^s}\Big[\sum_{k=0}^{\infty} \overline{\mathbf{c}}(x^f_k, x^s_k, a_k) \prod_{i=0}^{k-1} \alpha(x^f_i, x^s_i, a_i)\Big].$$

(b) *Using suitable transformations, the functional* (2.5) *can be also regarded as of an undiscounted type. Namely, including the dynamics in our model*

$$c_{k+1} = c_k + \mathbf{c}(x_k^f, x_k^s, a_k) d_k, \qquad d_{k+1} = d_k \alpha(x_k^f, x_k^s, a_k)$$

where the only initial values of interest are $c_0 = 0$ *and* $d_0 = 1$*, it is easy to see that* J *becomes*

$$J(x^f, x^s, \nu) = \liminf_{k\to\infty} E[c_k]$$

or without using c*,*

$$J(x^f, x^s, \nu) = E\Big[\sum_{k=0}^{\infty} \mathbf{c}(x_k^f, x_k^s, a_k) d_k\Big].$$

We now impose the following conditions.

Assumption 2.4. **(i)** *The cost function* $\mathbf{c}$ *is in* $\mathbb{B}^+(\mathbb{K})$*. There exists a constant* $\ell_0 > 0$ *with* $\ell(x^f, x^s, a) \geq \ell_0$ *for all* $(x^f, x^s, a) \in \mathbb{K} \cap (X \times V^s)$*.*

(ii) *The discount factor* α *belongs to* $\mathbb{B}^+(\mathbb{K})$*, and for every* $(x^f, x^s) \in X$ *it satisfies:*

$$\begin{aligned} &\alpha(x^f, x^s, a) = 1, \text{ when } a \in V^s \\ &\alpha(x^f, x^s, a) \leq \alpha_0 < 1, \text{ when } a \in V^f, \end{aligned} \tag{2.8}$$

for some given constant $0 < \alpha_0 < 1$*.*

Remark 2.5. **(a)** *Observe that the discount factor applied in* (2.5) *to the cost* $\mathbf{c}(x_k^f, x_k^s, a_k)$ *is*

$$\prod_{i=0}^{k-1} \alpha(x_i^f, x_i^s, a_i).$$

Assumption 2.4(ii) implies that this discounting equals the product of the $\alpha(x_i^f, x_i^s, a_i)$ *corresponding to actions* $a_i \in V^f$*, and so it incorporates exactly* $\mathtt{t}_k$ *factors which are less than or equal to* α_0*. This is exactly the discounting corresponding to the "natural time"* $\mathtt{t}_k$*. Now, if* $a_k \in V^f$ *then a further discounting* $\alpha(x_k^f, x_k^s, a_k) \leq \alpha_0$ *is applied to the forthcoming costs, while if* $a_k \in V^s$ *then no further discounting is applied at the present step:* $\alpha(x_k^f, x_k^s, a_k) = 1$*. This is consistent with the interpretation of hybrid control models: transitions of the special sub-dynamic occur* instantaneously*, and so in this case there is no further discounting. Besides, each transition according to the special sub-dynamic implies a* positive cost*, assumed to be bounded away from* 0*. In contrast, when the standard sub-dynamic is used, we indeed apply a discount factor, which is less than or equal to* $\alpha_0 < 1$*.*

(b) *Intuitively, the condition in Assumption 2.4(i) that the cost function* ℓ *is positive and bounded away from zero is intended to make that explosive sample paths have an associated total discounted which is infinite. Indeed, given an explosive sample path* $\omega \in H_\infty$ *there is some* k_0 *such that* $a_k \in V^s$ *for all* $k \geq k_0$*. In particular, for all* $k \geq k_0$ *we have*

$$\mathbf{c}(x_k^f, x_k^s, a_k) \prod_{i=0}^{k-1} \alpha(x_i^f, x_i^s, a_i) = \ell(x_k^f, x_k^s, a_k) \prod_{i=0}^{k_0-1} \alpha(x_i^f, x_i^s, a_i) \geq \ell_0 \prod_{i=0}^{k_0-1} \alpha(x_i^f, x_i^s, a_i).$$

Therefore, except for the somehow degenerate case that there would be some zero discountings, that is with $\alpha(x_i^f, x_i^s, a_i) = 0$*, the total discounted cost of this sample path would be*

$$\sum_{k=0}^{\infty} \mathbf{c}(x_k^f, x_k^s, a_k) \prod_{i=0}^{k-1} \alpha(x_i^f, x_i^s, a_i) = \infty.$$

Consequently, a finite expected discounted cost $J(x^f, x^s, \nu) < \infty$ *would imply that explosive sample paths would have* $P^{\nu}_{x^f,x^s}$*-probability zero.*

Our next assumption imposes finiteness of the optimal discounted cost function.

Assumption 2.6. *There exists a policy* $\nu \in \Pi$ *with* $J(x^f, x^s, \nu) < \infty$ *for all* $(x^f, x^s) \in X$.

Sufficient and necessary conditions for Assumption 2.6 are given in Propositions 2.7 and 2.8 below. Obviously, Assumption 2.6 implies that the optimal cost $J^*(x^f, x^s)$ is finite for any $(x^f, x^s) \in X$.

Sometimes we will need the stronger condition saying that there exists a policy $\nu \in \Pi$ such that $||J(\cdot, \nu)|| < \infty$, in which case we will have $||J^*|| < \infty$. This statement will be explicitly mentioned when needed.

The hybrid control model. Summarizing the previous paragraphs, a discrete-time Markov hybrid model can be seen as a tuple

$$\mathfrak{M} = (X^f \times X^s, V^f \cup V^s, \mathbb{K}, Q^f, Q^s, \alpha, c, \ell),$$

whose elements have been defined throughout this section.

Conditions for finiteness of J We conclude this section with some useful facts. Our next result gives a necessary condition for Assumption 2.6.

Proposition 2.7. *Suppose that Assumption 2.4 holds. If* $(x^f, x^s) \in X$ *and* $\nu \in \Pi$ *are such that* $J(x^f, x^s, \nu)$ *is finite, then*

$$\lim_{k\to\infty} E^{\nu}_{x^f,x^s}\Big[\prod_{i=0}^{k} \alpha(x^f_i, x^s_i, a_i)\Big] = 0.$$

Proof. To simplify the notation, we will write $x_i = (x^f_i, x^s_i) \in X$ for $i \geq 0$, with the initial state being $x_0 = (x^f_0, x^s_0) = (x^f, x^s)$.

On the set H_∞ of all histories, consider the following measurable sets

$$A_k = \{(x_0, a_0, x_1, a_1, \ldots) \in H_\infty : a_k \in V^s\} \quad \text{for } k \geq 0,$$

$$B = \{(x_0, a_0, x_1, a_1, \ldots) \in H_\infty : \alpha(x_k, a_k) > 0 \text{ for all } k \geq 0\}.$$

Choose a history

$$(x_0, a_0, x_1, a_1, \ldots) \in B \cap \liminf A_k.$$

There is some k_0 such that $a_k \in V^s$ for all $k \geq k_0$ and so, from Assumption 2.4, we have $\alpha(x_k, a_k) = 1$ and $\mathbf{c}(x_k, a_k) = \ell(x_k, a_k) \geq \ell_0$ for all $k \geq k_0$. Therefore, since $(x_0, a_0, \ldots) \in B$, for every $k \geq k_0$

$$\prod_{i=0}^{k-1} \alpha(x_i, a_i) = \prod_{i=0}^{k_0-1} \alpha(x_i, a_i) > 0,$$

and so

$$\sum_{k=0}^{\infty} \mathbf{c}(x_k, a_k) \prod_{i=0}^{k-1} \alpha(x_i, a_i) = \infty.$$

This yields that the total discounted cost of the sample paths in $B \cap \liminf A_k$ is infinite. The total expected discounted cost of ν being finite (this is precisely the hypothesis of this proposition), we must necessarily have $P_x^\nu\big(B \cap \liminf A_k\big) = 0$ or, equivalently,

$$P_x^\nu\big(B^c \cup \limsup A_k^c\big) = 1.$$

Now, take a history $(x_0, a_0, \ldots) \in B^c$. We have that $\prod_{i=0}^k \alpha(x_i, a_i)$ converges to 0 because at least one discount factor vanishes. On the other hand, take a history $(x_0, a_0, \ldots) \in \limsup A_k^c$. This means that $a_k \in V^f$ infinitely often, and so $\prod_{i=0}^k \alpha(x_i, a_i)$ also converges to 0 because $\alpha(x_k, a_k) \le \alpha_0$ infinitely often. Summarizing, we have shown that $\prod_{i=0}^k \alpha(x_i, a_i)$ converges with P_x^ν-probability one to 0. By dominated convergence, we conclude that

$$\lim_{k\to\infty} E_x^\nu\Big[\prod_{i=0}^k \alpha(x_i, a_i)\Big] = 0,$$

which completes the proof. □

We propose now a sufficient condition for Assumption 2.6. It uses the following notation. Given a transition probability measure $Q(\cdot|\cdot)$ on X given X, define $Q^1 = Q$ and recursively for $n \ge 1$

$$Q^{n+1}(B|x) = \int_X Q(B|y) Q^n(dy|x) \quad \text{for } B \in \mathcal{B}(X) \text{ and } x \in X,$$

which are the successive compositions of Q with itself. We will also use the following notation. Given $\mathtt{f} \in \mathbb{F}$, the kernel on X given X defined by $\mathbf{Q}(B|x, \mathtt{f}(x))$ for $x \in X$ and $B \in \mathcal{B}(X)$ (recall (2.1)) will be denoted by $\mathbf{Q}(\cdot|\cdot, \mathtt{f})$.

Proposition 2.8. *Suppose that Assumption 2.4 is satisfied and also that there exist* $\mathtt{f} \in \mathbb{F}$, $n \ge 1$, *and* $\epsilon > 0$ *such that, for all* $x = (x^f, x^s) \in X$,

$$\mathbf{Q}^n(D|x, \mathtt{f}) \le 1 - \epsilon,$$

where $D = \{x \in X : \mathtt{f}(x) \in V^s\}$. *Then there exists* $m_0 > 0$ *such that* $0 \le J^*(x) \le J(x, \mathtt{f}) \le m_0$ *for all* $x \in X$. *In particular, Assumption 2.6 holds.*

Proof. We will prove first the next preliminary result: Suppose that the initial state x is in D. Define $T = \min\{k : x_k \notin D\}$ as the exit time from D. Let us show that

$$P_x^{\mathtt{f}}\{T > kn\} \le (1-\epsilon)^k \quad \text{for all } k \ge 0. \tag{2.9}$$

Obviously, the inequality is true for $k = 0$, and for $k = 1$ because

$$P_x^{\mathtt{f}}\{T > n\} = P_x^{\mathtt{f}}\{x_1, x_2, \ldots, x_n \in D\} \le P_x^{\mathtt{f}}\{x_n \in D\} = \mathbf{Q}^n(D|x, \mathtt{f}) \le 1 - \epsilon. \tag{2.10}$$

Suppose now (2.9) holds for some $k \ge 1$. We have

$$\begin{aligned}
&P_x^{\mathtt{f}}\{T > (k+1)n\} = P_x^{\mathtt{f}}\{x_1, x_2, \ldots, x_{(k+1)n} \in D\} \\
&= E_x^{\mathtt{f}}\big[E_x^{\mathtt{f}}[\mathbf{1}_{\{x_1, x_2, \ldots, x_{(k+1)n} \in D\}} \mid h_{kn}]\big] \\
&= E_x^{\mathtt{f}}\big[\mathbf{1}_{\{x_1, x_2, \ldots, x_{kn} \in D\}} E_x^{\mathtt{f}}[\mathbf{1}_{\{x_{kn+1}, \ldots, x_{(k+1)n} \in D\}} \mid h_{kn}]\big] \\
&\le E_x^{\mathtt{f}}[\mathbf{1}_{\{x_1, x_2, \ldots, x_{kn} \in D\}}(1-\epsilon)] \le (1-\epsilon)^{k+1},
\end{aligned}$$

where we use (2.10) to bound the inner conditional expectation. Therefore, for every $x \in D$,

$$E_x^{\mathtt{f}}[T] = \sum_{k=0}^{\infty} P_x^{\mathtt{f}}\{T > k\} \le n \sum_{k=0}^{\infty} (1-\epsilon)^k = n/\epsilon. \tag{2.11}$$

We proceed with the proof of the proposition. We define the random variables T_r and S_r as the successive exit times from D and D^c, respectively. More specifically, define $T_1 = T$ and $S_1 = \min\{k > T_1 : x_k \in D\}$ and, for $r \ge 2$,

$$T_r = \min\{k > S_{r-1} : x_k \notin D\} \quad \text{and} \quad S_r = \min\{k > T_r : x_k \in D\}. \tag{2.12}$$

When the initial state x is in D then $T_1 > 0$, while if the initial state $x \notin D$, then $T_1 = 0$. If for some r we have $S_r < \infty$, then the arguments at the beginning of this proposition show that T_{r+1} is finite almost surely. Note that we may have $S_r = \infty$ for some $r \ge 1$ if the state process does not return to D.

Assume first that the initial state x is in D and make the convention that $S_0 = 0$. For any $k \ge 1$ and any sample path $\omega \in H_\infty$, define $n_k(\omega)$ as the number of times the state process has been outside D during the first $k-1$ periods:

$$n_k(\omega) = \sum_{j=0}^{k-1} \mathbf{1}_{D^c}(x_j).$$

Since $\alpha(y, \mathtt{f}(y)) = 1$ when $y \in D$ and $\alpha(y, \mathtt{f}(y)) \le \alpha_0$ when $y \notin D$, observe that

$$\begin{aligned} J(x, \mathtt{f}) &\le ||\mathbf{c}|| \cdot E_x^{\mathtt{f}}\Big[\sum_{k=0}^{\infty} \alpha_0^{n_k}\Big] \\ &= ||\mathbf{c}|| + ||\mathbf{c}|| \sum_{r=1}^{\infty} \Big(E_x^{\mathtt{f}}\big[\sum_{S_{r-1} < k \le T_r} \alpha_0^{n_k}\big] + E_x^{\mathtt{f}}\big[\sum_{T_r < k \le S_r} \alpha_0^{n_k}\big]\Big). \end{aligned}$$

When $T_r < k \le S_r$ the terms n_k increase by one. Grouping all such terms (plus the leftmost term) we get the whole series $||c|| \sum \alpha_0^k$. When $S_{r-1} < k \le T_r$ the term n_k remains constant and equal to

$$n_k = \sum_{j=1}^{r-1} (S_j - T_j) \ge r - 1$$

and so $\alpha_0^{n_k} \le \alpha_0^{r-1}$. Hence,

$$J(x, \mathtt{f}) \le \frac{||\mathbf{c}||}{1-\alpha_0} + ||\mathbf{c}|| \sum_{r=1}^{\infty} \alpha_0^{r-1} E_x^{\mathtt{f}}[T_r - S_{r-1}].$$

By the Markov property, at time S_{r-1} the process is in D and so the expected exit time from D is bounded by n/ϵ (recall (2.11)), and therefore

$$J^*(x) \le J(x, \mathtt{f}) \le \frac{||\mathbf{c}|| \cdot (1 + n/\epsilon)}{1-\alpha_0}.$$

When the initial state $x \notin D$ then the same arguments may be used to show that the above inequality remains true. □

Observe that Proposition 2.8 proves a result much stronger than Assumption 2.6. It shows that there exists $\mathtt{f} \in \mathbb{F}$ with bounded discounted cost and, hence, the optimal discounted cost function J^* is bounded as well.

3 Optimality results

In this section we study the solution to the hybrid control problem defined in (2.6). Our approach is the well-known dynamic programming method.

Dynamic programming equations. Given a function $u \in \mathbb{M}^+(X)$ we define the dynamic programming operator $\mathcal{T}u$ on X as follows:

$$\mathcal{T}u(x) = \inf_{a \in A(x)} \left\{ \mathbf{c}(x,a) + \alpha(x,a) \int_X u(y) \mathbf{Q}(dy \,|\, x, a) \right\} \quad \text{for } x \in X.$$

Taking into account the nature of the hybrid control model, we can define two associated operators $\mathcal{M}u$ and $\mathcal{H}u$ on X as

$$\mathcal{M}u(x^f, x^s) = \inf_{a \in A(x^f, x^s) \cap V^s} \left\{ \ell(x^f, x^s, a) + \int_X u(y^f, y^s) Q^s(dy^f \times dy^s \,|\, x^f, x^s, a) \right\}, \tag{3.1}$$

and

$$\mathcal{H}u(x^f, x^s) = \inf_{a \in A(x^f, x^s) \cap V^f} \left\{ c(x^f, x^s, a) + \alpha(x^f, x^s, a) \int_{X^f} u(y^f, x^s) Q^f(dy^f | x^f, x^s, a) \right\}. \tag{3.2}$$

Therefore, the dynamic programming operator $\mathcal{T}$ can be written, for $u \in \mathbb{M}^+(X)$ and $(x^f, x^s) \in X$, as

$$\mathcal{T}u(x^f, x^s) = \begin{cases} \mathcal{M}u(x^f, x^s), & \text{if } (x^f, x^s) \in D^\wedge, \\ \min\left\{ \mathcal{M}u(x^f, x^s), \mathcal{H}u(x^f, x^s) \right\}, & \text{if } (x^f, x^s) \in D^\vee \setminus D^\wedge, \\ \mathcal{H}u(x^f, x^s), & \text{if } (x^f, x^s) \in X \setminus D^\vee. \end{cases}$$

We define the so-named *dynamic programming equation*

$$u(x^f, x^s) = \mathcal{T}u(x^f, x^s) \quad \text{for all } (x^f, x^s) \in X, \tag{3.3}$$

which will be written, in short, by $u = \mathcal{T}u$.

Our next results will establish that the optimal discounted cost function J^* —recall (2.6)— is a solution of the dynamic programming equation, and we will show how to obtain an optimal policy from the fixed point equation $J^* = \mathcal{T}J^*$.

In addition to Assumptions 2.4 and 2.6, we will impose additional conditions on the control model. Namely, we will consider two alternative settings of hypotheses. One of them, see Assumption 3.1 below, will impose among other conditions that the transition kernel $\mathbf{Q}$ is strongly continuous (also referred to as strong Feller) and so we will refer to this case as to the *strongly continuous* case. On the other hand, Assumption 3.9 below will suppose that the transition kernel $\mathbf{Q}$ is weakly continuous (also referred to as weak Feller) and it will be referred to as the *weakly continuous* case. The terms strong and weak should not mislead the reader since both conditions are of a different nature, and Assumption 3.1 does not imply, in general, Assumption 3.9.

3.1 The strongly continuous case

The condition below states the hypotheses for the "strongly continuous" case.

Assumption 3.1. *For each $(x^f, x^s) \in X$ we have:*

(i) *The action set $A(x^f, x^s)$ is compact.*

(ii) *Given a bounded and measurable function $u : X \to \mathbb{R}$, the function*

$$a \mapsto \int_X u(y^f, y^s)\mathbf{Q}(dy^f \times dy^s | x^f, x^s, a)$$

is continuous in $a \in A(x^f, x^s)$.

(iii) *The functions $a \mapsto \mathbf{c}(x^f, x^s, a)$ and $a \mapsto \alpha(x^f, x^s, a)$ are lower semicontinuous on $A(x^f, x^s)$.*

We begin with a basic property about lower semicontinuity. For further details, see for instance, Proposition B.1 in Puterman [21].

Lemma 3.2. *Let Y be a complete and separable metric space, and $u : Y \to (0, \infty)$ and $v : Y \to [0, \infty]$ be lower semicontinuous functions. Then the sum $u + v$ and product $u \cdot v$ are lower semicontinuous functions on Y.*

In the sequel, to simplify the notation, the states of the system will be denoted simply by $x = (x^f, x^s) \in X$. The notation $dy = dy^f \times dy^s$ will be used as well.

Lemma 3.3. *Under Assumptions 3.1(ii)–(iii), given $u \in \mathbb{M}^+(X)$ and $x \in X$, the function $a \mapsto \mathbf{c}(x, a) + \alpha(x, a) \int_X u(y)\mathbf{Q}(dy|x, a)$ is lower semicontinuous on $A(x)$.*

Proof. Fix $u \in \mathbb{M}^+(X)$ and $x \in X$. We first prove that $a \mapsto \int_X u(y)\mathbf{Q}(dy|x, a)$ is lower semicontinuous on $A(x)$. Given $a \in A(x)$, suppose that the sequence $\{a_n\}$ in $A(x)$ converges to a. For all $k \geq 1$ define the function $u_k = \min\{k, u\}$. For each $k \geq 1$ we have

$$\liminf_{n\to\infty} \int_X u(y)\mathbf{Q}(dy|x, a_n) \geq \liminf_{n\to\infty} \int_X u_k(y)\mathbf{Q}(dy|x, a_n) = \int_X u_k(y)\mathbf{Q}(dy|x, a),$$

where the last equality follows from Assumption 3.1(ii) because u_k is bounded. Since this holds for all $k \geq 1$, by monotone convergence we get

$$\liminf_{n\to\infty} \int_X u(y)\mathbf{Q}(dy|x, a_n) \geq \int_X u(y)\mathbf{Q}(dy|x, a),$$

thus proving lower semicontinuity. The stated result now follows from Lemma 3.2. □

Note that, in this lemma, we are not excluding the possibility that

$$\int_X u(y)\mathbf{Q}(dy|x, a)$$

is infinite. In what follows, suppose that Assumption 3.1 holds. For $u \in \mathbb{M}^+(X)$ we recall that $\mathcal{T}u : X \to [0, \infty]$ is defined as

$$\mathcal{T}u(x) = \min_{a\in A(x)} \Big\{\mathbf{c}(x, a) + \alpha(x, a) \int_X u(y)\mathbf{Q}(dy|x, a)\Big\} \quad \text{for every } x \in X. \tag{3.4}$$

The fact that the minimum is attained follows because we are minimizing a lower semicontinuous function (Lemma 3.3) on the compact set $A(x)$ (Assumption 3.1(i)). Once again, we are not excluding the possibility that $\mathcal{T}u(x)$ is infinite. Observe that the operator $\mathcal{T}$ is monotone, meaning that given $u, v \in \mathbb{M}^+(X)$ such that $u \leq v$, we have $\mathcal{T}u \leq \mathcal{T}v$.

Our next result is a consequence of Proposition D.5 in [14] or Corollary 4.3 in [23].

Lemma 3.4. *Under Assumption 3.1, for every $u \in \mathbb{M}^+(X)$ the function $\mathcal{T}u$ is measurable and there exists $\mathtt{f} \in \mathbb{F}$ such that*

$$\begin{aligned}\mathcal{T}u(x) &= \min_{a\in A(x)} \Big\{ \mathbf{c}(x,a) + \alpha(x,a)\int_X u(y)\mathbf{Q}(dy|x,a) \Big\} \\ &= \mathbf{c}(x,\mathtt{f}(x)) + \alpha(x,\mathtt{f}(x))\int_X u(y)\mathbf{Q}(dy|x,\mathtt{f})\end{aligned}$$

for each $x \in X$.

Now we define recursively the functions u_k for $k \geq 0$. Let $u_0 := \mathbf{0}$ be the zero function on X and, for any $k \geq 0$, let $u_{k+1} := \mathcal{T}u_k$. Our next result explores some properties of the sequence $\{u_k\}_{k\geq 0}$.

Proposition 3.5. *Suppose that Assumptions 2.4, 2.6, and 3.1 hold.*

(i) *For each $k \geq 0$, the function u_k is in $\mathbb{B}^+(X)$.*

(ii) *The sequence $\{u_k\}$ converges monotonically to some $u^* \in \mathbb{M}^+(X)$ with $u^* \leq J^*$.*

(iii) *The function u^* satisfies $u^* = \mathcal{T}u^*$.*

Proof. (i). If $u_k \in \mathbb{B}^+(X)$ for some $k \geq 0$, then it is easily seen that $||\mathcal{T}u_k|| \leq ||\mathbf{c}|| + ||u_k||$. It follows that each $u_{k+1} := \mathcal{T}u_k$ is nonnegative, measurable (Lemma 3.4), and bounded. Hence $u_{k+1} \in \mathbb{B}^+(X)$.

(ii). Due that the cost function $\mathbf{c}$ is nonnegative, we have $u_0 \leq u_1$. Assuming, inductively, that $u_{k-1} \leq u_k$, by monotonicity of $\mathcal{T}$ we obtain $u_k \leq u_{k+1}$. Therefore, $\{u_k\}_{k\geq 0}$ converges monotonically to some nonnegative and measurable u^*. It remains to show that $u^*(x)$ is finite for every $x \in X$.

For every $k \geq 0$ define $\mathtt{f}_k \in \mathbb{F}$ as the measurable selector (recall Lemma 3.4) such that

$$u_{k+1}(x) = \mathbf{c}(x,\mathtt{f}_k(x)) + \alpha(x,\mathtt{f}_k(x))\int_X u_k(y)\mathbf{Q}(dy|x,\mathtt{f}_k) \quad \text{for } x \in X. \tag{3.5}$$

Now fix a $N \in \mathbb{N}$. Since we will be only concerned with the decision epochs $\{0,\dots,N\}$, we can assume without loss of generality that policies $\nu \in \Pi$ are restricted to the corresponding time horizon; hence, we will let $\nu = \{\nu_0,\dots,\nu_N\} \in \Pi$. Fix now $0 \leq t \leq N$ and $\nu \in \Pi$ and, for each history $h_t = (x_0, a_0, \dots, x_t) \in H_t$ up to time t, define

$$J_{t,N}(h_t,\nu) = E^{\nu}_{x_0}\Big[\sum_{k=t}^{N} \mathbf{c}(x_k,a_k)\prod_{j=t}^{k-1}\alpha(x_j,a_j) \mid h_t\Big].$$

Notice that $J_{t,N}(h_t,\nu)$ depends on ν only through the decision made at times $t,\dots,N$, that is, on $\{\nu_t,\dots,\nu_N\}$. Also, let

$$J_{t,N}(h_t) = \inf_{\nu\in\Pi} J_{t,N}(h_t,\nu) \quad \text{for } h_t \in H_t.$$

Define the policy $\nu^*_N = \{\mathtt{f}_N,\dots,\mathtt{f}_0\} \in \Pi$ on the time horizon $0,\dots,N$, with $\mathtt{f}_i$ $(i = 0,\cdots N)$ obtained as in (3.5). Our goal now is to show that, for fixed $N \geq 0$, we have

$$J_{t,N}(h_t) = J_{t,N}(h_t,\nu^*_N) = u_{N+1-t}(x_t) \quad \text{for every } 0 \leq t \leq N \text{ and } h_t \in H_t. \tag{3.6}$$

We will prove it by backwards induction on t. This equality is obvious for $t = N$ because for every $\nu \in \Pi$ and $h_N \in H_N$ we have

$$J_{N,N}(h_N, \nu) = \int_{A(x_N)} \mathbf{c}(x_N, a)\nu_N(da|h_N),$$

and so

$$J_{N,N}(h_N) = \min_{a\in A(x_N)} \mathbf{c}(x_N, a) = \mathbf{c}(x_N, f_0(x_N)) = J_{N,N}(h_N, \nu_N^*) = u_1(x_N).$$

Suppose that (3.6) holds for some $t + 1$ and let us prove it for t. Given arbitrary $\nu \in \Pi$ and $h_t \in H_t$ we have that

$$\begin{aligned}
J_{t,N}(h_t, \nu) &= E^{\nu}_{x_0}\Big[E^{\nu}_{x_0}\Big[\sum_{k=t}^{N} \mathbf{c}(x_k, a_k)\prod_{j=t}^{k-1} \alpha(x_j, a_j) \mid h_{t+1}\Big] \mid h_t\Big] \\
&= E^{\nu}_{x_0}\Big[\mathbf{c}(x_t, a_t) + \alpha(x_t, a_t)E^{\nu}_{x_0}\Big[\sum_{k=t+1}^{N} \mathbf{c}(x_k, a_k)\prod_{j=t+1}^{k-1} \alpha(x_j, a_j) \mid h_{t+1}\Big] \mid h_t\Big] \\
&= E^{\nu}_{x_0}\Big[\mathbf{c}(x_t, a_t) + \alpha(x_t, a_t)J_{t+1,N}(\nu, h_{t+1}) \mid h_t\Big] \\
&\geq E^{\nu}_{x_0}\Big[\mathbf{c}(x_t, a_t) + \alpha(x_t, a_t)u_{N-t}(x_{t+1}) \mid h_t\Big] \\
&= \int_{A(x_t)} [\mathbf{c}(x_t, a) + \alpha(x_t, a)\int_X u_{N-t}(y)\mathbf{Q}(dy|x_t, a)]\nu_t(da|h_t) \\
&\geq u_{N+1-t}(x_t),
\end{aligned}$$

with equality when $\nu = \nu_N^*$. This completes the backward induction argument. Hence, letting $t = 0$ (recall (3.6)) we have thus proved that for every $N \geq 0$ and $x \in X$

$$J_{0,N}(x) = J_{0,N}(x_0, \nu_N^*) = u_{N+1}(x).$$

Proceeding with the proof, the non negativity of the cost function implies that, for every $x \in X$, $N \geq 0$, and $\nu \in \Pi$,

$$u_{N+1}(x) \leq J_{0,N}(x, \nu) \leq J(x, \nu),$$

and so $u_{N+1}(x) \leq J^*(x)$ which, by Assumption 2.6, is finite. This shows that the limit function $u^* \leq J^*$ is finite on X, and we conclude that $u^* \in \mathbb{M}^+(X)$, which completes the proof of (ii).

(iii). Now that we know that $u^* \in \mathbb{M}^+(X)$, it turns out that $\mathcal{T}u^*$ is well defined. For all $k \geq 0$ we have $u_{k+1} = \mathcal{T}u_k \leq \mathcal{T}u^*$, by monotonicity of $\mathcal{T}$. Therefore, $u^* \leq \mathcal{T}u^*$.

To prove the reverse inequality, recall the definition of $\mathtt{f}_k \in \mathbb{F}$ given in (3.5). Fix $x \in X$ and, $A(x)$ being compact (Assumption 3.1(i)), there exists a subsequence $\{k_n\}_{n\geq 0}$ with $f_{k_n}(x) = a_{k_n} \to a^*$ for some $a^* \in A(x)$. Fix some n_0 and suppose that $n \geq n_0$. We have the following inequality

$$\int_X u_{k_n}(y)\mathbf{Q}(dy|x, a_{k_n}) \geq \int_X u_{k_{n_0}}(y)\mathbf{Q}(dy|x, a_{k_n}).$$

Take now the lim inf as $n \to \infty$ and use Assumption 3.1(ii) together with the fact that $u_{k_{n_0}} \in \mathbb{B}^+(X)$, to obtain

$$\liminf_{n\to\infty} \int_X u_{k_n}(y)\mathbf{Q}(dy|x, a_{k_n}) \geq \int_X u_{k_{n_0}}(y)\mathbf{Q}(dy|x, a^*).$$

But n_0 being arbitrary, monotone convergence yields

$$\liminf_{n\to\infty} \int_X u_{k_n}(y)\mathbf{Q}(dy|x, a_{k_n}) \geq \int_X u^*(y)\mathbf{Q}(dy|x, a^*). \tag{3.7}$$

Now, by lower semicontinuity of c and α (Assumption 3.1(iii)), and using (3.7), we obtain

$$\begin{aligned} u^*(x) &= \liminf_n u_{k_n+1}(x) \\ &\geq \liminf_n \mathbf{c}(x, a_{k_n}) + \liminf_n \Big[\alpha(x, a_{k_n}) \int_X u_{k_n}(y)\mathbf{Q}(dy|x, a_{k_n})\Big] \\ &\geq \mathbf{c}(x, a^*) + \alpha(x, a^*) \int_X u^*(y)\mathbf{Q}(dy|x, a^*) \;\geq\; \mathcal{T}u^*(x). \end{aligned}$$

This completes the proof that $u^* = \mathcal{T}u^*$. □

Hence, Proposition 3.5 shows that the operator $\mathcal{T}$ indeed has a fixed point in $\mathbb{M}^+(X)$. In Theorem 3.7 below we will see that, in fact, u^* equals J^*, the optimal discounted cost. We make the following comments regarding the inequality proved in (3.7).

Remark 3.6. *The inequality (3.7) is similar to the inequality given in the extension of Fatou's lemma in [15]. In our context, however, we cannot use that results because the norm of the u_k is not bounded in k. Here, in the proof of (3.7) we take advantage of the fact that we are dealing with nonnegative functions, and so we obtain the extended Fatou lemma dropping the uniformly bounded condition*

Theorem 3.7. *Suppose that Assumptions 2.4, 2.6, and 3.1 hold.*

(i) *The optimal discounted cost function J^* equals the limiting function u^* obtained in Proposition 3.5, and it is the minimal solution in $\mathbb{M}^+(X)$ of the dynamic programming optimality equation* (3.3).

(ii) *Any $\mathtt{f} \in \mathbb{F}$ attaining the minimum in the equation $J^* = \mathcal{T}J^*$, that is,*

$$J^*(x) = \mathbf{c}(x, \mathtt{f}(x)) + \alpha(x, \mathtt{f}(x)) \int_X J^*(y)\mathbf{Q}(dy|x, \mathtt{f}) \quad \textit{for all } x \in X$$

is optimal, and such $\mathtt{f} \in \mathbb{F}$ indeed exists.

(iii) *If, in addition, $J^* \in \mathbb{B}^+(X)$ then J^* is the unique solution in $\mathbb{B}^+(X)$ of the dynamic programming optimality equation.*

Proof. (i). Let $v \in \mathbb{M}^+(X)$ be a solution of $v = \mathcal{T}v$. By Lemma 3.4, there exists $\mathtt{f} \in \mathbb{F}$ with

$$v(x) = \mathbf{c}(x, \mathtt{f}(x)) + \alpha(x, \mathtt{f}(x)) \int_X v(y)\mathbf{Q}(dy|x, \mathtt{f}) \quad \text{for every } x \in X.$$

Iterating this equation we obtain that, for every $N \geq 0$ and $x \in X$

$$\begin{aligned} v(x) &= E_x^{\mathtt{f}}\Big[\sum_{k=0}^{N} \mathbf{c}(x_k, \mathtt{f}(x_k)) \prod_{i=0}^{k-1} \alpha(x_i, \mathtt{f}(x_i))\Big] + E_x^{\mathtt{f}}\Big[v(x_{N+1}) \prod_{i=0}^{N} \alpha(x_i, \mathtt{f}(x_i))\Big] \\ &\geq E_x^{\mathtt{f}}\Big[\sum_{k=0}^{N} \mathbf{c}(x_k, \mathtt{f}(x_k)) \prod_{i=0}^{k-1} \alpha(x_i, \mathtt{f}(x_i))\Big]. \end{aligned}$$

Letting $N \to \infty$ and using monotone convergence shows that that

$$v(x) \geq J(x, \mathtt{f}) \geq J^*(x). \tag{3.8}$$

We have thus established that if $v \in \mathbb{M}^+(X)$ is a solution of the dynamic programming optimality equation then $v \geq J^*$. On the other hand, in Proposition 3.5 we showed that $u^* \in \mathbb{M}^+(X)$ is a solution of the optimality equation with $u^* \leq J^*$. Thus $u^* = J^*$, which indeed solves the optimality equation and it is its minimal solution in $\mathbb{M}^+(X)$. This completes the proof of part (i).

(ii). To prove this statement, repeat the proof of (i) for $v = J^*$ to obtain (3.8). We necessarily have that f is optimal.

(iii). Suppose that $v \in \mathbb{B}^+(X)$ solves $v = \mathcal{T}v$. We know that, necessarily, $v \geq J^*$. Let us now show that $v \leq J^*$. Fix $x \in X$ and $\nu \in \Pi$. Obviously, if $J(x,\nu) = \infty$ we have $v(x) \leq J(x,\nu)$. Hence, in what follows we will suppose that $J(x,\nu) < \infty$. Using the inequality

$$v(x) \leq \mathbf{c}(x,a) + \alpha(x,a) \int_X v(y)\mathbf{Q}(dy|x,a) \quad \text{for all } (x,a) \in \mathbb{K},$$

for all $n \geq 0$ we have

$$\begin{aligned} E_x^\nu\Big[v(x_{n+1}) \prod_{i=0}^{n} \alpha(x_i,a_i) \mid h_n, a_n\Big] &= \prod_{i=0}^{n} \alpha(x_i,a_i) \int_X v(y)\mathbf{Q}(dy|x_n,a_n) \\ &= \prod_{i=0}^{n-1} \alpha(x_i,a_i) \cdot \Big[\mathbf{c}(x_n,a_n) + \alpha(x_n,a_n) \int_X v(y)\mathbf{Q}(dy|x_n,a_n) - \mathbf{c}(x_n,a_n)\Big] \\ &\geq \prod_{i=0}^{n-1} \alpha(x_i,a_i) \cdot \Big[v(x_n) - \mathbf{c}(x_n,a_n)\Big]. \end{aligned}$$

This implies, taking E_x^ν-expectation and rearranging terms,

$$E_x^\nu[\mathbf{c}(x_n,a_n) \prod_{i=0}^{n-1} \alpha(x_i,a_i)] \geq E_x^\nu[v(x_n) \prod_{i=0}^{n-1} \alpha(x_i,a_i)] - E_x^\nu[v(x_{n+1}) \prod_{i=0}^{n} \alpha(x_i,a_i)].$$

Summing these inequalities for $n = 0, 1, \ldots, N$ gives

$$E_x^\nu\Big[\sum_{n=0}^{N} \mathbf{c}(x_n,a_n) \prod_{i=0}^{n-1} \alpha(x_i,a_i)\Big] \geq v(x) - E_x^\nu\Big[v(x_{N+1}) \prod_{i=0}^{N} \alpha(x_i,a_i)\Big].$$

By monotone convergence, the lefthand term converges to $J(x,\nu)$, whereas the rightmost term converges to 0 as a consequence of Proposition 2.7 and the fact that v is bounded. Therefore, $J(x,\nu) \geq v(x)$, yielding that $J^* \geq v$. Hence, J^* is the unique solution of $v = \mathcal{T}v$ in $\mathbb{B}^+(X)$. □

We recall that a sufficient condition for $J^* \in \mathbb{B}^+(X)$ was given in Proposition 2.8.

Remark 3.8. *We have assumed that the running cost function* $\mathbf{c}$ *is nonnegative. This hypothesis is used, particularly, in Lemma 3.3 to ensure that application of the dynamic programming operator yields lower semi-continuous functions. This is based directly on Lemma 3.2, which indeed needs to deal with nonnegative functions.*

Typically, in the theory of discounted Markov decision processes with a constant *discount factor* $0 \leq \alpha_0 < 1$, *it is straightforward to generalize the dynamic programming results from a nonnegative running cost function to a bounded below running cost function* $\mathbf{c} \geq -M$ *for some* $M > 0$. *Namely, one considers the nonnegative cost function* $\mathbf{c} + M \geq 0$ *and then transforms the MDP with bounded below running cost function* $\mathbf{c}$ *into an equivalent MDP with nonnegative cost function:*

$$E_x^\nu\Big[\sum_k \alpha_0^k(\mathbf{c}(x_k,a_k) + M)\Big] = E_x^\nu\Big[\sum_k \alpha_0^k \mathbf{c}(x_k,a_k)\Big] + \frac{M}{1-\alpha_0},$$

and minimization of both expectations is equivalent.

In our context, however, such an approach to extend our hypotheses to the case of a bounded below cost function c (the running cost function under the standard sub-dynamic) is not possible. Indeed,

$$E^{\nu}_{x^f,x^s}\Big[\sum_{k=0}^{\infty}(\mathbf{c}(x^f_k,x^s_k,a_k)+M\mathbf{I}_{V^f}(a_k))\prod_{i=0}^{k-1}\alpha(x^f_i,x^s_i,a_i)\Big]$$
$$=E^{\nu}_{x^f,x^s}\Big[\sum_{k=0}^{\infty}\mathbf{c}(x^f_k,x^s_k,a_k)\prod_{i=0}^{k-1}\alpha(x^f_i,x^s_i,a_i)\Big]+M\cdot\sum_{k=0}^{\infty}E^{\nu}_{x^f,x^s}\Big[\mathbf{I}_{V^f}(a_k)\prod_{i=0}^{k-1}\alpha(x^f_i,x^s_i,a_i)\Big],$$

and note that the rightmost term is not constant *and depends on the policy used by the controller. This is because the discount factor is not constant and depends on the history of the state-action process. Hence, the problems with bounded below and nonnegative running cost function c are not necessarily equivalent.*

3.2 The weakly continuous case

In the "weakly continuous" setting we will impose the next condition.

Assumption 3.9. (i) *The multifunction $x\mapsto A(x)$ is compact-valued and upper semicontinuous.*

(ii) *Given bounded and continuous $u:X\to\mathbb{R}$, the function*

$$(x,a)\mapsto\int_X u(y)\mathbf{Q}(dy|x,a)$$

is continuous on $\mathbb{K}$

(iii) *The functions $\mathbf{c}$ and α are lower semicontinuous on $\mathbb{K}$.*

The proofs in this section somehow mimic the proofs in Section 3.1 and we will skip some details.

Lemma 3.10. *Under Assumptions 3.9(ii)–(iii), for every $u\in\mathbb{L}^+(X)$ the function $(x,a)\mapsto\mathbf{c}(x,a)+\alpha(x,a)\int_X u(y)\mathbf{Q}(dy|x,a)$ is lower semicontinuous on $\mathbb{K}$.*

Proof. The function u is the pointwise limit of a monotone nondecreasing sequence of nonnegative, bounded and continuous functions $u_n\uparrow u$ (see, e.g., Theorem 3.13 in [3]). Given $(x,a)\in\mathbb{K}$, suppose that the sequence (x_n,a_n) in $\mathbb{K}$ converges to (x,a). For each $k\geq 1$ we have

$$\liminf_{n\to\infty}\int_X u(y)\mathbf{Q}(dy|x_n,a_n)\geq\liminf_{n\to\infty}\int_X u_k(y)\mathbf{Q}(dy|x_n,a_n)=\int_X u_k(y)\mathbf{Q}(dy|x,a),$$

where the last equality follows from Assumption 3.9(ii) because u_k is bounded and continuous. By monotone convergence we get

$$\liminf_{n\to\infty}\int_X u(y)\mathbf{Q}(dy|x_n,a_n)\geq\int_X u(y)\mathbf{Q}(dy|x,a).$$

This proves that $(x,a)\mapsto\int_X u(y)\mathbf{Q}(dy|x,a)$ is lower semicontinuous on $\mathbb{K}$. The rest of the proof is similar to that of Lemma 3.3. □

At this point, recall the definition of the operator $\mathcal{T}$ given in (3.4). Our next follows from Proposition D.6 in [14] or Lemma 17.30 in [3].

Lemma 3.11. *Under Assumption 3.9, for every $u \in \mathbb{L}^+(X)$ the function $\mathcal{T}u$ is lower semicontinuous and there exists $\mathtt{f} \in \mathbb{F}$ such that*

$$\begin{aligned}\mathcal{T}u(x) &= \min_{a \in A(x)} \Big\{ \mathbf{c}(x,a) + \alpha(x,a) \int_X u(y)\mathbf{Q}(dy|x,a) \Big\} \\ &= \mathbf{c}(x,\mathtt{f}(x)) + \alpha(x,\mathtt{f}(x)) \int_X u(y)\mathbf{Q}(dy|x,\mathtt{f})\end{aligned}$$

for each $x \in X$.

Recall that in Section 3.1 we defined $u_0 := \mathbf{0}$ on X and $u_{k+1} := \mathcal{T}u_k$ for any $k \geq 0$.

Proposition 3.12. *Suppose that Assumptions 2.4, 2.6, and 3.9 hold.*

(i) *For each $k \geq 0$, the function u_k is in $\mathbb{L}^+(X)$.*

(ii) *The sequence $\{u_k\}$ converges monotonically to some $u^* \in \mathbb{L}^+(X)$ with $u^* \leq J^*$.*

(iii) *The function u^* satisfies $u^* = \mathcal{T}u^*$.*

Proof. (i). For every $k \geq 0$ we have that u_k is bounded and so, by Lemma 3.11 and the fact of that $\mathbf{c} \geq 0$, $u_{k+1} := \mathcal{T}u_k$ becomes nonegative, lower semicontinuous, and finite on X. That is, we have $u_{k+1} \in \mathbb{L}^+(X)$.

(ii). The sequence $\{u_k\}$ is monotone nondecreasing and so it converges to some lower semicontinuous function u^* on X. Proceeding as in the proof of Proposition 3.5(ii), we obtain that $u^* \leq J^*$, and so we indeed have $u^* \in \mathbb{L}^+(X)$.

(iii). For all $k \geq 0$ we have $u_{k+1} = \mathcal{T}u_k \leq \mathcal{T}u^*$, and so $u^* \leq \mathcal{T}u^*$.

For the reverse inequality, let $\mathtt{f}_k \in \mathbb{F}$ for $k \geq 0$ be such that (recall Lemma 3.11)

$$u_{k+1}(x) = \mathbf{c}(x,\mathtt{f}_k(x)) + \alpha(x,\mathtt{f}_k(x)) \int_X u_k(y)\mathbf{Q}(dy|x,\mathtt{f}_k) \quad \text{for } x \in X.$$

Fix arbitrary $x \in X$. Since $A(x)$ is compact, there exists a subsequence $\{k_n\}_{n \geq 0}$ with $f_{k_n}(x) = a_{k_n} \to a^*$ for some $a^* \in A(x)$. Fix some n_0 and suppose that $n \geq n_0$. We have the following inequality

$$\int_X u_{k_n}(y)\mathbf{Q}(dy|x,a_{k_n}) \geq \int_X u_{k_{n_0}}(y)\mathbf{Q}(dy|x,a_{k_n}).$$

The function $u_{k_{n_0}}$ is in $\mathbb{L}^+(X)$ and so, as in the proof of Lemma 3.10, we have that $\int_X u_{k_{n_0}}(y)\mathbf{Q}(dy|\cdot,\cdot)$ is lower semicontinuous on $\mathbb{K}$. Hence,

$$\liminf_{n\to\infty} \int_X u_{k_n}(y)\mathbf{Q}(dy|x,a_{k_n}) \geq \int_X u_{k_{n_0}}(y)\mathbf{Q}(dy|x,a^*).$$

Use now monotone convergence as $n_0 \to \infty$ to show that

$$\liminf_{n\to\infty} \int_X u_{k_n}(y)\mathbf{Q}(dy|x,a_{k_n}) \geq \int_X u^*(y)\mathbf{Q}(dy|x,a^*).$$

The rest of the proof is similar to that of Proposition □

The proof of our main result in this section is similar to that of Theorem 3.7 and we omit it.

Theorem 3.13. *Suppose that Assumptions 2.4, 2.6, and 3.9 hold.*

(i) *The optimal discounted cost function J^* equals the limiting function u^* obtained in Proposition 3.12, and it is the minimal solution in $\mathbb{L}^+(X)$ of the dynamic programming optimality equation* (3.3).

(ii) *There exists* $\mathtt{f} \in \mathbb{F}$ *attaining the minimum in the optimality equation* $J^* = \mathcal{T}J^*$*, and this policy is optimal.*

(iii) *If, in addition,* $J^* \in \mathbb{B}^+(X)$ *then* J^* *is the unique solution in* $\mathbb{L}^+(X) \cap \mathbb{B}^+(X)$ *of the dynamic programming optimality equation.*

The same comments as in Remark 3.8 on the possibility to extend Theorem 3.13 to a bounded below cost function c are in order.

3.3 Contact set and continuation region

Our previous results, Theorems 3.7 and 3.13, give sufficient conditions ensuring that the optimal discounted cost function (2.6) is a solution of the dynamic programming equation (3.3), and show how to obtain an optimal deterministic stationary policy from the fixed point equation $J^* = \mathcal{T}J^*$. With the notation introduced in (3.1) and (3.2), by letting $x := (x^f, x^s)$, equation (3.3) reads

$$J^*(x) = \begin{cases} \mathcal{M}J^*(x), & \text{if } x \in D^\wedge, \\ \min\big\{\mathcal{M}J^*(x), \mathcal{H}J^*(x)\big\}, & \text{if } x \in D^\vee \setminus D^\wedge, \\ \mathcal{H}J^*(x), & \text{if } x \in X \setminus D^\vee. \end{cases}$$

Moreover, let $\mathtt{f} \in \mathbb{F}$ be such that it attains the minimum in the fixed point equation $J^* = \mathcal{T}J^*$. We have that $\mathtt{f}$ is an optimal policy and, besides, given a state $x \in D^\vee \setminus D^\wedge$, we have the following situations:

$$\begin{aligned} &\bullet \text{ if } \mathcal{M}J^*(x) > \mathcal{H}J^*(x) \text{ then the optimal action } \mathtt{f}(x) \text{ is in } V^f; \\ &\bullet \text{ If } \mathcal{M}J^*(x) < \mathcal{H}J^*(x) \text{ then the optimal action } \mathtt{f}(x) \text{ is in } V^s; \\ &\bullet \text{ If } \mathcal{M}J^*(x) = \mathcal{H}J^*(x) \text{ then the optimal action } \mathtt{f}(x) \text{ can be taken either in } V^s \\ &\qquad\qquad \text{or in } V^f. \end{aligned} \tag{3.9}$$

Obviously, when the state is in $D^\wedge$ or in $X \setminus D^\vee$, then optimal actions are necessarily in V^s or in V^f, respectively.

We introduce the set D^* defined as:

$$D^* = \big\{x \in D^\vee : J^*(x) = \mathcal{M}J^*(x)\big\},$$

This set is so-named *contact set* and it can be regarded as an *optimal* region, in the sense that outside its closure, the optimal choice is to apply a fast action in V^f, and as a consequence the standard sub-dynamic is activated, and once the state of the system reaches the interior of D^*, an optimal rule is to apply an action in V^s and thus the special sub-dynamic is turned on. This fact can be summarized as follows.

Proposition 3.14. *An optimal rule outside* $\bar{D}^*$ *must be necessarily on* V^f*, while in* $\mathring{D}^*$*, the optimal rule can be taken on* V^s*.*

Proof. First note that the admissibility condition of control policies on the region $X \setminus D^\vee$, is to apply an action $a \in V^f$. Also, the dynamic programming equation (3.3) shows that when $x \in D^\vee$ but $x \notin \bar{D}^*$, we have

$$J^*(x) = \min\big\{\mathcal{M}J^*(x), \mathcal{H}J^*(x)\big\}. \tag{3.10}$$

and necessarily the next cases hold: (a) $J^*(x) > \mathcal{M}J^*(x)$ or (b) $J^*(x) < \mathcal{M}J^*(x)$. We will show that the optimal rule should be to apply an action in V^f. Indeed, by (3.10) and considering case (a) first, we deduce

$\min\{\mathcal{M}J^*(x), \mathcal{H}J^*(x)\} > \mathcal{M}J^*(x)$, which produces a contradiction. On the other hand, combining case (b) with (3.10), we obtain $\min\{\mathcal{M}J^*(x), \mathcal{H}J^*(x)\} < \mathcal{M}J^*(x)$, which gives $\mathcal{H}J^*(x) < \mathcal{M}J^*(x)$. Then, in virtue of (3.9), the optimal rule is to apply the control $\mathtt{f}(x) \in V^f$.

Likewise, if $x \in \mathring{D}^*$, we have again two cases: (a) $x \in D^\wedge$ or (b) $x \in D^\vee \setminus D^\wedge$. Case (a) directly follows from the definition of admissible policies. Now, for case (b), we have that equation (3.10) holds true and, on $\mathring{D}^*$, we deduce that

$$\min\{\mathcal{M}J^*(x), \mathcal{H}J^*(x)\} = \mathcal{M}J^*(x) \leq \mathcal{H}J^*(x).$$

Then, using again (3.9), we see that an optimal rule must necessarily satisfy $\mathtt{f}(x) \in V^s$ when $\mathcal{M}J^*(x) < \mathcal{H}J^*(x)$ or to (optionally) choose as optimal rule $\mathtt{f}(x) \in V^s$ when $\mathcal{M}J^*(x) = \mathcal{H}J^*(x)$. □

The previous rule on how to apply optimal actions is clear when the state is located on the regions inside and outside D^*; however, our present hypotheses are insufficient to see what happens on the boundary ∂D^* of D^*. A sufficient condition that extends Proposition 3.14 over all X is to assume that the set D^* is closed (or even open); another sufficient condition can be given in the spirit of the continuity of J^* and $\mathcal{M}J^*$. For instance, we can assert the following result:

Proposition 3.15. *Assume that J^* and $\mathcal{M}J^*$ are continuous functions on X. Then, the contact set D^* is closed, implying that the optimal action on ∂D^* must be taken in V^s.*

Proof. Take a sequence $\{x_k\}$ in D^* so that $x_k \to x$, for some $x \in X$. Then, for each k, we have

$$J^*(x_k) = \mathcal{M}J^*(x_k), \quad k \geq 1.$$

Taking lim in both sides of the above expression, and using the continuity of J^* and $\mathcal{M}J^*$, we obtain $J^*(x) = \mathcal{M}J^*(x)$, which implies that D^* is closed.

The fact that the optimal action on ∂D^* must be taken in the set V^s, follows by applying the same arguments of Proposition 3.14. □

The well-posedness of the contact set D^* allows us to define the so-named *continuation region* that is very common in impulsive and switching control problems. The continuation region is simply the complement of the contact set D^* and, as the name suggests, the optimal rule signalizes to *continue* applying the standard sub-dynamic as long as the state is located outside the contact set D^*.

Time-interface set. The time-interface set consists of real positive numbers that signalize the times when the special sub-dynamic must be activated in an optimal way. It can be obtained in terms of the contact set D^* as follows: let

$$\omega = (x_0, a_0, \cdots, x_k, a_k, \cdots)$$

be an element of $H_\infty = \mathbb{K}^\infty$, with $x_k := (x_k^f, x_k^s)$, we recursively define:

$$\mathtt{k}_i \equiv \mathtt{k}(\omega, i) = \inf\{k \geq \mathtt{k}(\omega, i-1) + 1 : J^*(x_k) = \mathcal{M}J^*(x_k)\} \quad \forall\ i \geq 0, \tag{3.11}$$

with $\mathtt{k}(\omega, -1) = -1$, which is regarded as a stopping time relative to the history, i.e., $\mathtt{k}(\omega, i)$ is a random variable with values in $\{0, 1, 2, \cdots, \infty\}$. These random variables must satisfy the following conditions:

$$\begin{aligned}
&\mathtt{k}^\wedge(\omega, i) \leq \mathtt{k}(\omega, i) \leq \mathtt{k}^\vee(\omega, i), \quad \forall i \geq 0, \text{ with} \\
&\mathtt{k}^\wedge(\omega, i) = \inf\{k \geq i : (x_k^f, x_k^s, a_k) \in D^\vee \times V^{\mathtt{s}}\}, \\
&\mathtt{k}^\vee(\omega, i) = \inf\{k \geq i : (x_k^f, x_k^s, a_k) \in D^\wedge \times V^{\mathtt{s}}\}.
\end{aligned}$$

Remark 3.16. *Observe that the elements of the time-interface-set can be also interpreted in terms of the first exit times from the continuation region of the process $\{x_k\}$.*

4 Examples

This section is devoted to show simple applications that illustrate our results herein. We want to clarify that these examples are presented only for illustrative purposes, so we will not explicitly develop here all the previous optimality results.

A consumption-investment problem with market modes Suppose that a small agent (or investor) wishes to allocate his investment among various assets with different return rates and, in turn, he is able to change his investment into different market modes at some specific times. To simplify our model, we will consider only two market modes, denoted by m_1 and m_2. We shall assume that the agent can invest his wealth in only two assets: a risk-free asset (e.g. a bond) with a fixed interest rate denoted by $r(m_j)$, $j = 1, 2$ and a risky asset (e.g. stock or commodity) with a stochastic return rate $\xi_k(m_j)$, $j = 1, 2$ at time k. We shall assume that, for each $j = 1, 2$, the random variables $\{\xi_k(m_j)\}_k$ are independent and identically distributed with common distribution μ_j. Note that these rates are different depending on market mode the agent is trading. We define a consumption-investment strategy denoted by $\nu = \{(\mathsf{c}_k, i_k),\ k \geq 0\}$ representing the investment (portfolio) process $\{i_k\}$ and the consumption process $\{\mathsf{c}_k\}$. In this case, at any time k, i_k represents the fraction of wealth invested in the risky stock, $(1 - i_k)$ is the portion invested in the risk-free asset, and c_k the amount of wealth to be consumed. We assume that

$$0 \leq i_k \leq 1, \quad 0 \leq \mathsf{c}_k \leq x_k^f \qquad \forall\ k \geq 0,$$

where x_k^f denotes the investor's wealth at time k. Assuming that the market is self-financing, a suitable model for the investor's wealth is

$$x_{k+1}^f = \left[(1 - i_k)(1 + r(x_k^s)) + i_k \xi_k(x_k^s)\right](x_k^f - \mathsf{c}_k) \quad k \geq 0,$$

with a given initial wealth $x_0^f = x > 0$, initial mode $x_0^f = m_i$ (for some fixed $i = 1, 2$). In this last equation, the (slow) variable x_k^s means the market mode at time $k \geq 0$, with values in $X^s := \{m_1, m_2\}$. It can be verified (use Example C.7 in Hernández-Lerma & Lasserre [14]) that the transition kernels Q^f and Q^s defined in (2.1) become weakly continuous.

A typical choice of reward c is a given utility function $u(\cdot)$ dependent of the investor's consumption c. That is:

$$c(x^f, x^s, \mathsf{c}, i) := u(\mathsf{c}).$$

Now, we take into account the fact that the investor can decide at some times $\tau_k \in \mathbb{N}$ if his investment is subject to change from a mode m_1 to m_2 and vice versa. Each mode change is supposed to be instantaneous in time and it may generate a *positive* penalty ℓ, that may also depend on the configuration to be selected (e.g. $\ell(m_1, m_2)$ or $\ell(m_2, m_1)$). We can assume that such a penalty is bounded away from zero by some constant that may represent a predetermined *minimum fee* for changing between market modes.

The control model is given as follows: The state space $X = \mathbb{R}_+ \times \{m_1, m_2\}$, the action space $A := \mathbb{R}_+ \times [0, 1] \cup \{m_1, m_2\}$, and the admissible action space $A(x^f, x^s) := [0, x^f] \times [0, 1] \cup \{m_1, m_2\}$. Usually, the investor has no restrictions to change of configuration at any moment, then we may assume $D^\vee = X$ and $D^\wedge = \emptyset$. The times τ_k can be regarded as stopping times representing the instants when the action a is chosen on $X^s = \{m_1, m_2\}$. We also have that the cost rate becomes

$$\mathbf{c}(x^f, x^s, a) = u(a)\mathbf{1}_{\{\mathbb{R}_+ \times [0,1]\}}(a) + \ell(x^s, a)\mathbf{1}_{\{m_1, m_2\}}(a).$$

and the joint wealth (dynamic) is given by

$$\underbrace{\left(x^f_{k+1}, x^s_{k+1}\right) = \left(\left[(1-i_k)(1+r(x^s_k)) + i_k\xi_k(x^s_k)\right](x^f_k - \mathsf{c}_k), x^s_k\right)}_{\text{standard sub-dynamic}} \quad \text{if} \ \ a_k \in [0, x^f_k] \times [0,1],$$

or

$$\underbrace{\left(x^f_{k+1}, x^s_{k+1}\right) = \left(x^f_k, [x^s_k]^{\mathsf{c}}\right)}_{\text{special sub-dynamic}} \quad \text{if} \ \ a_k \in \{m_1, m_2\},$$

where $[x^s_k]^{\mathsf{c}}$ represents the complement of x^s_k.

Using the payoff function (2.5) as our optimality criterion, the control problem is then is to find an optimal consumption-investment-configuration strategy $a \in \mathbb{R}_+ \times [0, x^f] \cup \{m_1, m_2\}$ that maximizes the total discounted wealth (2.5).

Remark 4.1. **(i)** *Observe that this optimal control model can be regarded as a composed control problem formed by a Markov decision process*[8] *(MDP) together with a switching control model.*

(ii) *The context of this example leads to* maximize *a given reward or revenue function. This problem can be easily posed in a minimization context; in this sense, our present theory does apply.*

(iii) *The assumptions imposed in this example do not contradict the hypotheses given in previous sections. In fact, we open the possibility to work with a specific model like this in order to obtain specific optimal consumption-invested control policies over different market modes.*

A manufacturing-production problem Consider a manufacturing-production system in which a given company produces a single item. The production is made by means of m machines, each of them has two different modes: in operation or closed, represented by the quantities 1 and 0, respectively. The state variables x^f_k will represent the inventory of the items at time $k \geq 0$ with values in $\mathbb{N}$, whereas x^s_k is the state of the machines configurations with values in the set $X^s = \{(a_1, \cdots, a_m) : a_i \in \{0,1\}, \ i = 1, \cdots, m\}$. At each period of time k, the control variable a_k can be the quantity produced (and immediately supplied) by the company or either a quantity to be bought to external competitors or the change of a different machine mode. Assuming a finite storage capacity, say, C, the action and admissible action sets become $A = [0, C] \cup [0, C] \cup X^s$ and $A(x^f, x^s) = [0, C - x^f] \cup [0, C - x^f] \cup X^s$, respectively. The reason why we are repeating the set $[0, C]$ (resp. $[0, C - x^f]$) is to distinguish the items to be produced by the company and the items that are bought to the competitors. These two sets are assumed to be different from each other.

On the other hand, let us allow negative inventory levels by assuming that excess of demand is backlogged and filled when additional inventory enters the company. Then, the changes of the inventory and machine configurations can be modeled by means of the following dynamic.

[8] or conventional control model

$$\underbrace{(x^f_{k+1}, x^s_{k+1}) = (x^f_k + a_k - \xi_k, x^s_k)}_{\text{standard sub-dynamic}} \quad \text{if} \ \ a_k \in \big[0, C - x^f_k\big],$$

or

$$\underbrace{(x^f_{k+1}, x^s_{k+1}) = (a_k + \eta_k, x^s_k)}_{\text{special sub-dynamic}} \quad \text{if} \ \ a_k \in \big[0, C - x^f_k\big],$$

or

$$\underbrace{(x^f_{k+1}, x^s_{k+1}) = (x^f_k, G(x^s_k))}_{\text{special sub-dynamic}} \quad \text{if} \ \ a_k \in X^s,$$

where $\{\xi_k\}_k$ and $\{\eta_k\}_k$ are two independent sequence of i.i.d. random variables representing the demand of the product and the shipping loses or defective pieces when the purchase is carried out, each of them defined on $\mathbb{R}$. Actually, we may assume that such sequences have densities ρ_ξ and ρ_η that are continuous on $\mathbb{R}$. Then, according to Example C8 in Hernández-Lerma & Lasserre [14], the transition kernels Q^f and Q^s given in (2.1) are strongly continuous, provided that the function $G : X^s \mapsto X^s$ is continuous as well.

Finally, suppose that we want to minimize an expected operator cost (2.5), in which the elements of $\mathbf{c}(\cdot)$ are defined as follows: we have an inventory-production cost c given by

$$c(x^f, x^s, a) := c_1 a + c_2 \max(0, x^f) + c_3 \max(0, -x^f),$$

where c_1 is the unit production cost, c_2 the unit holding cost for excess of inventory and c_3 the penalization cost for the unfilled demand. Furthermore, we have another cost

$$\ell(x^f, x^s, a) = \ell_1(a)\mathbf{1}_{\{a\in[0,C-x^f]\}} + \ell_2(x^s, a)\mathbf{1}_{\{a\in X^s\}},$$

with elements $\ell_1(a) := c_4 a$ representing the cost of purchasing a items at the price $c_4 > 0$, and $\ell_2(x^s, a)$ a given continuous function representing the cost of changes between configuration machines.

The optimal (hybrid) control problem is then to find an optimal control policy a^* that consists in a series of production, purchases and configuration decisions in such a way that it minimizes a payoff (2.5).

As a last comment, it is worth mentioning that the two costs ℓ_1 and ℓ_2, were attached to a single cost ℓ because they are involved in the special dynamic, whose transitions are instantaneous in time.

Remark 4.2. **(i)** *This example becomes a composed control problem, formed by a conventional control model MDP, an impulsive control problem, and a switching control problem; all of them in a unified model.*

(ii) *Similarly to the previous example, the assumptions imposed in this example do not contradict the hypotheses given in previous sections. In fact, we open a way to study this specific model by using the theoretical results provided in previous pages.*

5 Special cases

In this section we present some important particular cases of hybrid systems.

Impulse control A simple case of a hybrid model is the so-called *optimal impulse control problem.* In this model, the dynamic system follows its "natural" random evolution and the controller selects the times at which he acts on the dynamic system. This model can be interpreted as a Markov chain with controlled discontinuities. This type of models is comprehensively studied in Bensoussan [4], Bensoussan & Lions [5, 6], Menaldi [19], Robin [25], Stettner [26, 27], among others.

Impulse-type models become special classes of hybrid models studied in the past sections. Indeed, consider X^s as a singleton, so that we may identify the state space X with X^f only. We shall assume that X^f is a subset of a vectorial space with sum and scalar product well defined on it (for instance $X^f \subset \mathbb{R}$).

As for the available actions for the controller, we define $V^f = \{\Delta\}$, so that, $A(x) = A = \{\Delta\} \cup V^s \; \forall \, x \in X$, and finally, assume $D^\wedge = \emptyset$ and $D^\vee = X$. In this case Δ is interpreted as the absence of controller's standard actions.

The dynamic follows the rule

$$\mathbf{Q}(dy|x,a) = \begin{cases} Q^f(dy|x) & \text{if } a = \Delta \\ \delta_{\{x+a\}}(dy) & \text{if } a \in V^s \subset X^f, \end{cases}$$

which can be also expressed as

$$\underbrace{x_{k+1} = F(x_k, \Delta, w_k)}_{\textit{standard sub-dynamic}} \text{ if } a_k = \Delta \qquad \text{or} \qquad \underbrace{x_{k+1} = x_k + a_k}_{\textit{special sub-dynamic}} \quad \text{if } a_k \in V^s, \tag{5.1}$$

where $\{w_k\}$ is a sequence of i.i.d. random variables (recall Remark 2.1). The running cost and the discount factor functions are usually of the type

$$\mathbf{c}(x,a) = c(x)\mathbf{1}_{\{\Delta\}}(a) + \ell(x,a)\mathbf{1}_{V^s}(a), \; \gamma > 0 \qquad \text{and} \qquad \alpha(x,a) = \hat{\alpha}(x)\mathbf{1}_{\{\Delta\}}(a) + \mathbf{1}_{V^s}(a)$$

for some discount function $\hat{\alpha}$ on X, satisfying conditions similar to those in Assumptions 2.4 and 3.1 (or 3.9). The dynamic programming equation is

$$u(x) = \min\big\{\mathcal{M}u(x), \mathcal{H}u(x)\big\} \quad \text{for } x \in X,$$

where

$$\begin{aligned} \mathcal{H}u(x) &= c(x) + \hat{\alpha}(x) \int_X u(y) Q^f(dy|x) \\ \mathcal{M}u(x) &= \inf_{a \in V^s} \Big\{ \ell(x,a) + u(x+a) \Big\}. \end{aligned}$$

Switching control Another class of hybrid model is the *switching control problem*. The dynamic system can operate under several *modes* or *configurations*, and the controller decides the times when the dynamics switches from one mode to another one. Typically, running and switching costs are incurred. Here, the fast variable indicates the state of the system, whereas the slow variable gives the current mode. Related works in continuous-time models are, for instance, Bensoussan & Lions [5], Menaldi & Blankenship [20], Zhang et. al. [31] and the references therein.

In order to describe the switching problem in our context, we let:

- $V^s = X^s$, so that the mode is seen both as an action and a state (or label);
- $V^f = \{\Delta\}$ and $A = \{\Delta\} \cup X^s$, where as before, Δ denotes the absence of actions from the controller;
- $D^\wedge = \emptyset$ and $D^\vee = X^f \times X^s$.

The particular feature of the dynamic is that changing the mode does not affect the fast state. More precisely,

$$\mathbf{Q}(dy^f \times dy^s | x^f, x^s, a) = \begin{cases} Q^f(dy^f | x^f, x^s)\delta_{\{x^s\}}(dy^s) & \text{if } a = \Delta \\ \delta_{\{x^f\}}(dy^f) \cdot \delta_{\{a\}}(dy^s) & \text{if } a \in V^s, \end{cases}$$

or it can be expressed in terms of an explicit dynamic by

$$\underbrace{(x^f_{k+1}, x^s_{k+1}) = \big(F(x^f_k, x^s_k, w_k), x^s_k\big)}_{\textit{standard sub-dynamic}}, \quad a_k \in \Delta \qquad \text{or} \qquad \underbrace{(x^f_{k+1}, x^s_{k+1}) = \big(x^f_k, a_k\big)}_{\textit{special sub-dynamic}}, \quad a_k \in V^s = X^s,$$

where, as before, $\{w_k\}$ represents a sequence of i.i.d. random variables.

The one-stage cost function $\mathbf{c}(\cdot)$ incorporates the running cost c and the switching cost ℓ:

$$\mathbf{c}(x^f, x^s, a) = c(x^f, x^s)\mathbf{1}_{\{\Delta\}}(a) + \ell(x^f, x^s, a)\mathbf{1}_{V^s}(a),$$

whereas the discount factor is given by

$$\alpha(x^f, x^s, a) = \hat{\alpha}(x^f, x^s)\mathbf{1}_{\{\Delta\}}(a) + \mathbf{1}_{V^s}(a),$$

for some function $0 < \hat{\alpha} < 1$ satisfying Assumptions 2.4 and 3.1 (or 3.9).

The dynamic programming equation, in this case, is given by:

$$u(x^f, x^s) = \min\big\{\mathcal{M}u(x^f, x^s), \mathcal{H}u(x^f, x^s)\big\} \quad \text{for } (x^f, x^s) \in X,$$

where $\mathcal{H}$ and $\mathcal{M}$ satisfy

$$\begin{aligned} \mathcal{H}u(x^f) &:= c(x^f, x^s) + \hat{\alpha}(x^f, x^s) \int_{X^f} u(y^f, x^s) Q^f(dy^f | x^f, x^s), \\ \mathcal{M}u(x^f) &:= \inf_{a \in X^s} \big\{\ell(x^f, x^s, a) + u(x^f, a)\big\}. \end{aligned}$$

Standard or conventional control problem It should be clear that a standard control problem with varying discount factor can also be analyzed with the techniques herein, just by only considering the regular sub-dynamic. This class of models have been exhaustively studied and, for discrete-time models, we can quote Bensoussan [4], Hernández-Lerma & Lasserre [14, 15], Puterman [21] and the references therein. To do so, we consider X^s to be a singleton, so that we identify X^f with the state space X, and we let V^s and $D^\vee$ to be empty. The dynamic system follows the stochastic kernel $Q^f(dy|x, a)$, or, equivalently, the dynamic

$$\underbrace{x_{k+1} = F(x_k, a_k, w_k)}_{\textit{standard sub-dynamic}} \quad a \in V^f,$$

where, as in previous cases, $\{w_k\}$ denotes a sequence of i.i.d. random variables.

As in the previous cases, c and α are the running and discount functions introduced in Section 2. In this case the dynamic programming equations is the standard one:

$$u(x) = \mathcal{H}u(x) \quad \text{for } x \in X^f$$

where

$$\mathcal{H}u(x) = \inf_{a \in A(x)} \big\{c(x, a) + \alpha(x, a) \int_X u(dy) Q^f(dy|x, a)\big\}.$$

References

[1] Abate, A., Amin, S., Prandini, M., Lygeros, J., Sastry, S.: (2006). Reachability Analysis for Controlled Discrete-Time Stochastic Hybrid Systems. In A. Hespanha, J., & Tiwari A. (Eds.), *Lecture notes in computer science,* **3927,** pp. 49–63. Springer-Verlag Berlin Heidelberg.

[2] Abate, A., Prandini, M., Lygeros, J., Sastry, S.: (2008). Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica*, **44**, 2724-2734.

[3] Aliprantis, C. D, Border, K.C.: *Infinite dimensional analysis.* Springer-Verlag, New York, 2006.

[4] Bensoussan, A.: *Dynamic programming and inventory control.* IOS Press, Amsterdam, 2011.

[5] Bensoussan A. and Lions, J.L.: *Applications of Variational Inequalities in Stochastic Control*, North-Holland, Amsterdam, 1982. (pages are referred to the French edition)

[6] Bensoussan, A.; Lions, J.L.: *Impulse Control and Quasi-Variational Inequalities.* Gauthier-Villars, Paris, 1984. (pages are referred to the French edition)

[7] Bensoussan, A., Menaldi J.L.: (1997) Hybrid control and dynamic programming. *Dynamics and continuous, discrete and impulsive systems* **3**, pp. 395–442.

[8] Bensoussan, A., Menaldi J.L.: (2000) Stochastic hybrid control *J. Math. Anal. Appl.* **249**, pp. 261–288.

[9] Bertsekas, D.P., Shreve, S.E.: *Stochastic Optimal Control: The Discrete-Time Case.*, Athena Scientific, Belmont, M.A, 1996.

[10] Goebel, R., Sanfelice, R.G., Teel, A.R.: *Hybrid Dynamical Systems: Modeling, Stability, and Robustness.* Princeton University Press, Princeton, N.J., 2012.

[11] Borkar, V.S. and Ghosh, M.K. and Sahay, P.: (1999) Optimal control of a stochastic hybrid system with discounted cost *J. Optim. Theory Appl.* **101**, pp. 557–580.

[12] Branicky, M. S.: Studies in hybrid systems: Modeling, analysis, and control, Thesis MIT, LIDS-TH-2304 (1995), Cambridge, USA.

[13] Branicky, M.S; Borkar, V.S; Mitter, S.K.: (1998) A unified framework of hybrid control: model and optimal control theory *IEEE Trans. Autom. Control* **43**, pp. 31–45.

[14] Hernández-Lerma, O., Lasserre, J.B.: *Discrete-time Markov control processes: Basic optimality criteria.* Springer-Verlag, New York, 1996.

[15] Hernández-Lerma, O., Lasserre, J.B.: *Further topics on discrete-time Markov control processes.* Springer-Verlag, New York, 1999.

[16] Jasso-Fuentes, H; Menaldi, J.L., Prieto-Rumeau, T., Robin, M.: (2018) Discrete-time hybrid control in Borel spaces: average cost optimality criterion. *J. Math. Anal. Appl.* **462**, pp. 1695–1713.

[17] Kallenberg, O. (2002). *Foundations of Modern Probability, 2nd edition.* Springer-Verlag, New York.

[18] Lygeros, J.: (2005). An Overview of Hybrid Systems Control. In A. Hristu-Varsakelis, D., & Levine W. S.. (Eds.), *Handbook of Networked and Embedded Control Systems,* **IV,** pp. 519–537. Birkhäuser, Boston.

[19] Menaldi J.L.: (1987) Optimal impulse control problems for degenerate diffusions with jumps, *Acta Appl. Math.* **8**, pp. 165–198.

[20] Menaldi J.L.; Blankenship, G.L.: (1984) Optimal stochastic scheduling of power generation systems with scheduling delays and large cost differentials, *SIAM J.Control Optim.*, **22**, pp. 121–132.

[21] Puterman, M.L.: *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons Inc, New York, 1994.

[22] Quadrat, J.P.: (1980) Existence de solution et algorithme de resolutions numériques de problémes stochastiques degenerées ou non, *SIAM J. Control and Optim.*, **18**, 199–226.

[23] Rieder. U.: (1978) Measurable selection theorems for optimization problems *Manuscripta Math.* **24**, pp. 115–131.

[24] Robin, M.: (1976) Contrle impulsionnel avec retard pour les processus de diffusion, C.R.A.S, **282** 463–466.

[25] Robin, M.: *Controle impulsionnel des processus de Markov*, Thesis INRIA, TE-035, Paris, France, 1977.

[26] Stettner, L.: (1986) Discrete Time Adaptive Impulsive Control Theory, *Stoch. Processes and their Appl*, **23**, 177–197.

[27] Stettner, L.: (1989) On some stopping and impulsive control problems with a general discount rate criterion, *Probab. Math. Statist.*, **10**, 223–245.

[28] Summers, S., Lygeros, J.: (2010) Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Automatica*, **46**, 1951-1961.

[29] Yin, G.G., Zhang, Q.: *Discrete-Time Markov Chains. Two-Time-Scale Methods and Applications* Springer, New York, 2005.

[30] Yin, G.G., Zhu, C.: *Hybrid Switching Diffusions* Springer, New York, 2010.

[31] Zhang, W., Jianghai, H., Abate, A.: (2012) Infinite-Horizon Switched LQR Problems in Discrete Time: A Suboptimal Algorithm With Performance Analysis, *IEEE Transactions on Automatic Control*, **57**, 1815–1821.