Kybernetika

Yofre H. García; Juan González-Hernández

Solutions of semi-Markov control models with recursive discount rates and approximation by ϵ-optimal policies

Kybernetika, Vol. 55 (2019), No. 3, 495–517

Persistent URL: `http://dml.cz/dmlcz/147862`

Terms of use:

© Institute of Information Theory and Automation AS CR, 2019

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

SOLUTIONS OF SEMI-MARKOV CONTROL MODELS WITH RECURSIVE DISCOUNT RATES AND APPROXIMATION BY ϵ–OPTIMAL POLICIES

Juan González-Hernández and Yofre H. García

This paper studies a class of discrete-time discounted semi-Markov control model on Borel spaces. We assume possibly unbounded costs and a non-stationary exponential form in the discount factor which depends of on a rate, called *the discount rate.* Given an initial discount rate the evolution in next steps depends on both the previous discount rate and the sojourn time of the system at the current state. The new results provided here are the existence and the approximation of optimal policies for this class of discounted Markov control model with non-stationary rates and the horizon is finite or infinite. Under regularity condition on sojourn time distributions and measurable selector conditions, we show the validity of the dynamic programming algorithm for the finite horizon case. By the convergence in finite steps to the value functions, we guarantee the existence of non-stationary optimal policies for the infinite horizon case and we approximate them using non-stationary $\epsilon-$optimal policies. We illustrated our results a discounted semi-Markov linear-quadratic model, when the evolution of the discount rate follows an appropriate type of stochastic differential equation.

Keywords: optimal stochastic control, dynamic programming method, semi-Markov processes

Classification: 93E20, 49L20

1. INTRODUCTION

The theory of discrete-time discounted semi-Markov control models (SMCMs) has been widely developed in many directions. The first assumptions to guarantee the existence of optimal policies in this type control models was initially formulated on finite spaces by [6], [15], [16], [20], later, on denumerable state spaces and compact control spaces in [5], [8] (also know as Markov chains), and generalized to Borel spaces by [3],[17]. Applications and other contributions can also be seen in [18], [19],[11], [13], [7], [14], and the references therein. However, there are very few SMCMs with a non-constant discount factor. A work with some similar conditions was proposed by Zagst in [23], where a particular separable Bayesian SMCM is considered on Borel spaces with variable discount factor that depends of the observation and the control. He proves the existence

DOI: 10.14736/kyb-2019-3-0495

of non-stationary optimal policies under the so-called Blackwell- sufficiency order of probability measures.

We consider here a discounted SMCM in discrete-time where the discount factor has an exponential form and depends on a non-stationary rate which is obtained recursively from an initial discount rate and is dependent of the control. We refer to this model as *discounted Semi-Markov Control Model with recursive discount rates.* The discount rates in each step are obtained by applying a different recursive function, depending on the previous discount rate and a non-negative random variable which represents the current sojourn time of the system. More precisely, if the initial state and the initial discount rate are x_0 and r_0, respectively, then recursive relation of the discount rate in this SMCM takes the form

$$r_{n+1} := E[R_n(r_n, \delta_{n+1}) \mid x_n, r_n, a_n], \quad n = 0, 1, 2, \ldots, \tag{1}$$

where E is the expectation respect to the random variable δ_{n+1} on $(0, \infty)$ that represents the sojourn time of the system state at step n conditioned by (x_n, r_n, a_n). The above dependence can be used in the modeling of semi-Markov control systems where the discount rate is oscillating around an ideal value. So this model be more realistic than the case when the discount factor is considered as constant; for example, to find optimal actions of big investors or monopolies.

The main contribution focuses in give conditions and prove on both the existence of non-stationary optimal policies as the possibility of approximation with $\epsilon-$optimal policies for this type of discounted case when the horizon is finite or infinite.

The rest of paper has been organized as follows. In Section 2 we introduce the components of the SMCM, the sets of histories and admissible policies, the regularity condition to guarantee that the process does not collapse in finite time, and the canonical construction. In Section 3, the value function and the optimal control problem with finite horizon are defined. Also, to guarantee the existence of non-stationary optimal policies for the SMCM, the inf-compactness, continuity respect to controls, and the regularity condition assumption are supposed. The existence of a non-stationary optimal policy for finite case is obtained from a version of the Dynamic Programming Theorem. In Section 4 we discussed the case of SMCM, of the finite horizon value functions to the discounted optimal value with infinite horizon. We also prove the non-stationary optimal policy can be approximated by non-stationary $\epsilon-$optimal policies. In Section 5, we give an example of a recursive discount rate whenever the discount rate evolution follows a stochastic differential equation (SDE). Finally, in Section 6, a semi-Markov linear-quadratic model is presented as an example.

2. CONTROL MODEL

Let X and Y be Borel spaces. A stochastic kernel on X given Y is a function $P(\cdot \mid \cdot)$ such that $P(\cdot \mid y)$ is a probability measure on X for each $y \in Y$, and $P(A \mid \cdot)$ is a measurable function on Y for each fixed $A \in \mathfrak{B}(X)$. The Borel sigma algebra of subsets of X, denoted by $\mathfrak{B}(X)$, is the smallest sigma algebra that contains the open sets of X. The function $f : X \to Y$ is measurable if for any open set B on Y, the inverse image $f^{-1}(B)$ is a Borel set in X. For any measurable set X, we denote by $\mathbb{P}(X)$ the collection of all probability measures over on X. $\mathbb{P}(X)$ is a Borel space when X is a Borel space (see for instance [2], or [4, chapter 7]).

The SMCM we deal with is

$$M = \Big(X', \{R_n \mid n = 0, 1, 2, \dots\}, A, \{A(x, r) \mid (x, r) \in X'\}, F, Q, D, d\Big), \tag{2}$$

where:

- The set $X' := X \times [d_1, d_2]$, where X is the Borel state space. The interval $[d_1, d_2]$, where $0 < d_1 \leq d_2$, is the discount rate space. The transitions of system state in the time occur in the steps $n = 0, 1, 2, \dots$ *The holding time of SMCM* at any state, between step n and step $n + 1$, is a random variable δ_n which takes values on $(0, \infty)$. In addition, $\delta_0 := 0$ and $T_n := \delta_0 + \dots + \delta_n$ represents the accumulated time up to step n, $n = 0, 1, 2, \dots$

- $\{R_n \mid n = 0, 1, 2, \dots\}$ is a sequence of measurable functions (called the *recursive discount rate functions*)

  $$R_n : [d_1, d_2] \times (0, \infty) \to [d_1, d_2],\ n = 0, 1, 2, \dots$$

  From which generates in each period the discount rate according to (1).

- A is a Borel space and represents the set of controls.

- The family $\{A(x, r) \mid (x, r) \in X'\}$ of non-empty measurable subsets of A, where $A(x, r)$ is the subset of admissible controls for $(x, r) \in X'$, such that,

  $$\mathbb{K} := \{(x, r, a) \mid a \in A(x, r), (x, r) \in X'\}, \tag{3}$$

  is a measurable subset $X' \times A$.

- The continuous function $t \to F(t \mid x, a, x')$ is a probability distribution function for each $(x, a, x') \in X \times A \times X$, and we assume that $F(t \mid \cdot)$ is jointly measurable for each real number t.

- The stochastic kernel Q on X given $\mathbb{K}$ is the transition law between states.

- The measurable function $D : X \times A \to \mathbb{R}$, denotes the immediate cost that the system incurs when the state is x_n and a control a_n is chosen. The measurable function $d : X \times A \to \mathbb{R}$ stands for the holding cost of the system at sate x_n when the discount rate is r_n and the control is a_n.

Form of F and Q we can define the distribution function of the sojourn time δ_{n+1} of system conditional to $(x_n, r_n, a_n) \in \mathbb{K}$ (see [22], Section 1.2) as

$$G(t \mid x_n, r_n, a_n) := P(\delta_{n+1} \leq t \mid x_n, r_n, a_n) = \int_X F(t \mid x_n, a_n, x')Q(dx' \mid x_n, r_n, a_n) \tag{4}$$

for $t \in \mathbb{R},\ n = 0,1,2,\dots,\ x' \in X$. To simplify the notation we will denote by $\overline{Q}_{n+1}$, $n = 0,1,2,\dots,$ the transitions between the pairs state-discount rate in each step that now can be write as

$$\overline{Q}_{n+1}(B \times F \mid x_n, r_n, a_n, \delta_{n+1}) := Q(B \mid x_n, r_n, a_n) I_F[R_n(r_n, \delta_{n+1})], \tag{5}$$

$B \in \mathfrak{B}(X),\ F \in \mathfrak{B}\big([d_1, d_2]\big)$, $n = 0,1,2,\dots,$ where $I_F(\cdot)$ is the indicator function on the set F.

Assumption 2.1. The set $\mathbb{K}$ contains the graph of a measurable function f from X' to A, such that, $f(x,r) \in A(x,r)$, for all $(x,r) \in X'$.

For the SMCM (2), the total holding cost function between step n and step $n+1$ is given by

$$c(x_n, r_n, a_n) = D(x_n, a_n) + d(x_n, a_n) \int_0^\infty \int_0^t e^{-r_n s} ds G(dt \mid x_n, r_n, a_n), \tag{6}$$

for $n = 0,1,2,\dots$

2.1. Interpretation

Suppose the initial state and the initial discount rate of the system are $x_0 = x$ and $r_0 = r$ respectively. The initial discount rate $r_0 \in [d_1, d_2]$ is given. The initial control $a_0 = a \in A(x_0, r_0)$ is chosen. Therefore, the system generates the immediate cost $D(x_0, a_0)$ at time T_0, and remains in the state $x_0 = x$ during a (non-negative) random time $\delta_1 := T_1 - T_0$ with distribution $G(\cdot \mid x, r, a)$. Additionally, the system generates the initial cost $d(x_0, a_0)$ of sojourn during the time interval $[T_0, T_1)$ in the state x_0 when control a_0 was selected. Then, the total holding cost function between step $n = 0$ and $n = 1$ is

$$c(x_0, r_0, a_0) = D(x_0, a_0) + d(x_0, a_0) \int_0^\infty \int_0^t e^{-r_0 s} ds G(dt \mid x_0, r_0, a_0), \tag{7}$$

and the system is moved to state $x_1 = y$ according to the distribution $Q(\cdot \mid x_0, r_0, a_0)$ at time $T_1 = T_0 + \delta_1$. The discount rate of step 1 is

$$r_1 = E[R_0(r_0, \delta_1) \mid x_0, r_0, a_0], \tag{8}$$

where R_0 is the discount rate recursive function. This evolution continues in the same way throughout the horizon.

2.2. Histories and Policies

The set of admissible histories at step n is denoted by H_n, and is defined in the usual way

$$H_0 := X', \quad H_n := \big(\mathbb{K} \times (0,\infty)\big)^n \times X', \quad n = 1,2,\dots, \tag{9}$$

where $\mathbb{K}$ is defined in (3). An $n-$admissible history $h_n \in H_n$ is a vector

$$(x_0, r_0, a_0, \delta_1, x_1, r_1, a_1, \delta_2, \ldots, x_{n-1}, r_{n-1}, a_{n-1}, \delta_n, x_n, r_n), \tag{10}$$

with $(x_i, r_i, a_i) \in \mathbb{K}$, for $i = 0, \ldots, n-1$, and $(x_n, r_n) \in X'$.

Definition 2.2. (Policies of SMCM) Let M be the SMCM (2).

(a) A policy is a sequence of stochastic kernels $\pi = \{\pi_n\}_{n\in\mathbb{N}}$ on A given H_n, such that,

$$\pi_n(A(x_n, r_n) \mid h_n) = 1, \quad \forall h_n \in H_n,\, n = 0, 1 \ldots$$

The set of all policies is denoted by Π.

(b) A policy π is a Markov policy if exists a sequence of stochastic kernels $\{\phi_n\}_{n\in\mathbb{N}}$ on A given X' that satisfies

$$\pi_n(D \mid h_n) = \phi_n(D \mid x_n, r_n),$$

for all $h_n \in H_n,\ D \in \mathfrak{B}(A)$, and $n \in \mathbb{N}$. The set of all Markov policies is denoted $\mathbb{M}$.

(c) A Markov policy π is non-stationary deterministic if exists a sequence of stochastic kernels $\{g_n\}_{n\in\mathbb{N}}$ of measurable functions (or *selectors*) $g_n : X' \to A$ such that

$$\phi_n(D \mid x_n, r_n) = I_D[g_n(x_n, r_n)],$$

for all $(x_n, r_n) \in X',\ D \in \mathfrak{B}(A)$, and $n \in \mathbb{N}$. The set of all deterministic policies is denoted $\mathbb{D}$.

Clearly $\mathbb{D} \subset \mathbb{M} \subset \Pi$.

2.3. The canonical construction of an SMCM

Let $(\Omega, \mathfrak{F})$ be the measurable space of all trajectories where $\Omega := (X' \times A \times (0, \infty))^\infty$ and $\mathfrak{F}$ is the respective σ-algebra of subsets of Ω. The subset $H_\infty := (\mathbb{K} \times (0, \infty))^\infty$ of Ω is the set of all admissible trajectories.

Given a policy $\pi = \{\pi_n\}$ and the initial pair $(x_0, r_0) \in X'$, the Ionescu Tulcea Theorem (see [4], Proposition 7.28) guarantees the existence of a probability measure P_ν^π on $(\Omega, \mathfrak{F})$ such that

$$\begin{aligned} &P_\nu^\pi(dx_0 dr_0 da_0 d\delta_1 dx_1 dr_1 da_1 d\delta_2 \cdots) \\ &\quad := \nu(dx_0, dr_0)\pi_0(da_0 \mid x_0, r_0) G(d\delta_1 \mid x_0, r_0, a_0) \\ &\qquad \cdot \overline{Q}_1(dx_1, dr_1 \mid x_0, r_0, a_0, \delta_1)\pi_1(da_1 \mid x_1, r_1) \cdots, \end{aligned}$$

where ν is an arbitrary initial distribution for (x_0, r_0). P_ν^π satisfies the following properties:

For all $B \in \mathfrak{B}(X),\ F \in \mathfrak{B}([d_1, d_2]),\ D \in \mathfrak{B}(A),\ t \in \mathbb{R}$, and $h_n \in H_n$, $n = 0, 1, \ldots$

$$P_\nu^\pi\big((x_0, r_0) \in B \times F\big) = \nu(B \times F), \tag{11}$$

$$P^\pi_\nu\big(x_{n+1}\in B \mid h_n, a_n\big) = Q(B \mid x_n, r_n, a_n), \tag{12}$$
$$P^\pi_\nu\big((x_{n+1}, r_{n+1})\in B\times F \mid h_n, a_n, \delta_{n+1}\big) = \overline{Q}_{n+1}(B\times F \mid x_n, r_n, a_n, \delta_{n+1}), \tag{13}$$
$$P^\pi_\nu\big(a_n \in D \mid h_n\big) = \pi_n(D \mid x_n, r_n), \tag{14}$$
$$P^\pi_\nu(\delta_{n+1}\le t \mid h_n, a_n) = G(t \mid x_n, r_n, a_n). \tag{15}$$

For each $\pi\in\Pi$ and $\nu\in\mathbb{P}(X')$ the stochastic processes $\{x_n, r_n, a_n, \delta_{n+1}\}$, $n = 0,1,\dots,$ defined on

$$(\Omega, \mathfrak{F}, P^\pi_\nu)$$

is called the semi-Markov control processes. Denote by $P^\pi_{(x,r)}$ when ν is concentrated at (x,r) and the conditional expectation operator associated with $P^\pi_{(x,r)}$ is denoted by $E^\pi_{(x,r)}$.

Assumption 2.3. (Regularity condition) There exist real numbers $\epsilon > 0$ and $\gamma > 0$ such that

$$G(\gamma \mid x, r, a) \le 1-\epsilon, \quad \text{for all } (x,r,a)\in X'\times A. \tag{16}$$

Definition 2.4. For each $(x,r,a)\in\mathbb{K}$, let us define

$$\Delta(x,r,a) := \int_0^\infty e^{-rt} G(dt \mid x, r, a) \tag{17}$$

and

$$\tau(x,r,a) := \frac{1-\Delta(x,r,a)}{r}. \tag{18}$$

From (17) and (18), the total sojourn cost function (6) can be rewritten as

$$c(x,r,a) = D(x,a) + \tau(x,r,a)d(x,a), \quad (x,r,a)\in\mathbb{K}. \tag{19}$$

Remark 2.5. Note that, by (15), is valid for all $t_n > 0$, $n = 0,1,2,\dots$

$$P\left(\delta_1 \le t_1, \delta_2\le t_2, \dots, \delta_{n+1}\le t_{n+1} \mid h_n\right) = \prod_{i=0}^n G\left(t_{i+1} \mid x_i, r_i, a_i\right). \tag{20}$$

Proposition 2.6. Suppose that the Assumption 2.3 holds. Then for any $n = 1,2,\dots$

(i) $\rho := \sup_{\mathbb{K}} \Delta(x,r,a) < 1.$

(ii) $P^\pi_{(x,r)}\left[\sum_{n=1}^\infty \delta_n = \infty\right] = 1, \quad \forall (x,r)\in X'.$

Proof. Let ϵ and γ satisfy Assumption 2.3.

(i) By Riemann-Stieltjes formula (integration by parts)

$$\int_{\alpha}^{\beta} f(t)dg(t) + \int_{\alpha}^{\beta} g(t)df(t) = f(\beta)g(\beta) - f(\alpha)g(\alpha), \tag{21}$$

and using the functions $f(t) = e^{-rt}$ and $g(t) = G(t \mid x, r, a)$, and the values $\beta = \infty,\ \alpha = 0$, we obtain

$$\int_{0}^{\infty} e^{-rt} G(dt \mid x, r, a) + \int_{0}^{\infty} G(t \mid x, r, a)(de^{-rt}) = 0.$$

Hence

$$\begin{aligned}\Delta(x, r, a) &= \int_{0}^{\infty} e^{-rt} G(dt \mid x, r, a)\\ &= r\int_{0}^{\infty} e^{-rt} G(t \mid x, r, a)dt\\ &= r\int_{0}^{\gamma} e^{-rt} G(t \mid x, r, a)dt\\ &\qquad + r\int_{\gamma}^{\infty} e^{-rt} G(t \mid x, r, a)dt.\end{aligned}$$

By the regularity condition 2.3, we obtain

$$\begin{aligned} r\int_{0}^{\gamma} e^{-rt} G(t \mid x, r, a)dt &\le r(1-\epsilon)\int_{0}^{\gamma} e^{-rt}dt \\ &\le (1-\epsilon)(1-e^{-r\gamma}),\end{aligned} \tag{22}$$

and since $G(t \mid x, r, a) \le 1$,

$$\begin{aligned} r\int_{\gamma}^{\infty} e^{-rt} G(t \mid x, r, a)dt &\le \int_{\gamma}^{\infty} e^{-rt}dt \\ &\le e^{-r\gamma},\end{aligned} \tag{23}$$

for all $(x, r, a) \in \mathbb{K}$. Then, combining (22) and (23) we have

$$\begin{aligned}\Delta(x, r, a) &\le (1-\epsilon)(1-e^{-r\gamma}) + e^{-r\gamma}\\ &= 1 - \epsilon(1-e^{-r\gamma})\\ &\le 1 - \epsilon(1-e^{-d_1\gamma})\end{aligned}$$

for every $(x, r, a) \in \mathbb{K}$, and hence

$$\rho := \sup_{\mathbb{K}} \Delta(x, r, a) \le 1 - \epsilon(1-e^{-d_1\gamma}) < 1.$$

(ii) Let $\pi \in \Pi$ and $(x_0, r_0, a_0, \delta_1, x_1, r_1, a_1, \delta_2, \dots) \in H_\infty$. Since that for all $n = 0, 1, 2, \dots$, the discount rate $r_n \in [d_1, d_2]$,

$$E^{\pi}_{(x,r)}\left[\prod_{n=0}^{\infty} \Delta(x_n, r_n, a_n)\right]$$

$$
\begin{aligned}
&= E^{\pi}_{(x,r)}\left[\prod_{n=0}^{\infty}\int_{0}^{\infty} e^{-r_n t} G(dt \mid x_n, r_n, a_n)\right] \\
&= E^{\pi}_{(x,r)}\left[\prod_{n=0}^{\infty} E^{\pi}_{(x,r)}[e^{-r_n \delta_{n+1}} \mid x_n, r_n, a_n]\right] \\
&\geq E^{\pi}_{(x,r)}\left[\prod_{n=0}^{\infty} E^{\pi}_{(x,r)}[e^{-d_2 \delta_{n+1}} \mid x_n, r_n, a_n]\right] \\
&= E^{\pi}_{(x,r)}\left[E^{\pi}_{(x,r)}\left[\exp\left\{-\sum_{n=0}^{\infty} d_2 \delta_{n+1}\right\} \mid x_0, r_0, a_0, x_1, r_1, a_1, \cdots\right]\right], \qquad (24)
\end{aligned}
$$

where the last equality is consequence of (20). On the other hand, by part (i), $\Delta(x,r,a) \leq \rho < 1$ for all $(x,r,a) \in \mathbb{K}$, then

$$\prod_{k=0}^{n} \Delta(x_k, r_k, a_k) \leq \rho^n, \ \forall n.$$

Thus, $\prod\limits_{k=0}^{n} \Delta(x_k, r_k, a_k)$ tends to zero as $n \to \infty$, and therefore

$$E^{\pi}_{(x,r)}\left[\prod_{n=0}^{\infty} \Delta(x_n, r_n, a_n)\right] = 0. \qquad (25)$$

By (24) and (25)

$$
\begin{aligned}
0 &\leq E^{\pi}_{(x,r)}\left[\exp\left\{-d_2 \sum_{n=1}^{\infty} \delta_n\right\} \mid x_0, r_0, a_0, x_1, r_1, a_1, \cdots\right], \\
&\leq E^{\pi}_{(x,r)}\left[\prod_{n=0}^{\infty} \Delta(x_n, r_n, a_n)\right] \\
&= 0,
\end{aligned}
$$

and by no-negativity of the exponential function

$$P^{\pi}_{(x,r)}\left[\sum_{n=1}^{\infty} \delta_n = \infty\right] = 1, \quad \forall (x,r) \in X', \quad \pi \in \Pi.$$

□

3. THE FINITE HORIZON SEMI-MARKOV CONTROL PROBLEM

Let us consider the SMCM in (2) and assume the admissible set of policies Π to be nonempty. Define the function $J : \Pi \times X' \to \mathbb{R}$ by

$$J(\pi, x, r) := E^{\pi}_{(x,r)}\left[\sum_{n=0}^{N} e^{-S_n} c(x_n, r_n, a_n)\right], \quad (x,r) \in X', \qquad (26)$$

where $S_0 := 0$, and

$$S_n := \sum_{i=0}^{n-1} r_i \delta_{i+1}, \; n = 1, 2, \ldots . \tag{27}$$

The measurable function J is called *the expected discounted total cost with recursive discount rate,* or simply, the discounted total index.

The expression

$$J^*(x, r) := \inf_{\Pi} \; J(\pi, x, r), \quad (x, r) \in X', \tag{28}$$

denotes the discounted optimal value of the system when the optimization horizon is N.

The *semi-Markov control problem* consists to finding the policy π^* such that

$$J(\pi^*, x, r) = J^*(x, r), \quad (x, r) \in X'.$$

Now we introduce the semicontinuous - semicompactness conditions (see [12] subsection 3.3) on the SMCM that guarantee the existence of measurable selectors and, furthermore, the existence of optimal policies:

Definition 3.1. A function $\nu : \mathbb{K} \to \mathbb{R}$ is called *inf-compact* on $\mathbb{K}$, if for each $(x, r) \in X'$ and $z \in \mathbb{R}$, the set

$$\{a \in A(x, r) : \nu(x, r, a) \le z\}$$

is compact.

Assumption 3.2.

1. For each $(x, r) \in X'$, the total sojourn cost c is such that $c(x, r, \cdot)$ is lower-semi continuous function on $A(x, r)$, bounded below and inf-compact on $\mathbb{K}$.

2. For each $n = 0, 1, 2, \ldots$ the transition law $\overline{Q}_{n+1}$ is strongly continuous, that is, the function

$$\bar{u}_{n+1}(x, r, a) := \int_{X'} u(y, r') \overline{Q}_{n+1}(dy, dr' \mid x, r, a) \tag{29}$$

is continuous and bounded on $\mathbb{K}$, for every measurable bounded function u on X'.

Note that by Assumption 3.2 and the continuity of F, the functions $\Delta(x, r, \cdot)$ and $\tau(x, r, \cdot)$ are continuous for all $(x, r) \in X'$.

Assumption 3.3. There exists a policy $\pi \in \Pi$ such that $J(\pi, x, r) < \infty$ for any $x \in X$ and r. The set of policies that satisfy this condition is denoted by Π^0.

3.1. Dynamic programming algorithm

The Semi-Markov control problem with finite horizon N is solved with the next version of the Dynamic Programming Theorem:

Theorem 3.4. Let, for $(x,r) \in X'$,

$$J_N(x,r) := \min_{a \in A(x,r)} c(x,r,a), \tag{30}$$

and for $n = 0, 1, \dots, N-1$,

$$J_n(x,r) := \min_{a \in A(x,r)} \Big[c(x,r,a) + \iint\limits_{X\,0}^{\quad\infty} e^{-rt} J_{n+1}(y, R_n(r,t)) G(dt \mid x,r,a) \cdot Q(dy \mid x,r,a) \Big], \tag{31}$$

for $(x,r) \in X'$. Under Assumptions 3.2, the functions J_n are measurable for each $n = 0, 1, 2, \dots, N$, and exist measurable selectors $f_n \in \mathbb{D}$, such that

$$J_n(x,r) = c(x,r,f_n) + \iint\limits_{X\,0}^{\quad\infty} e^{-rt} J_{n+1}(y, R_n(r,t)) \times G(dt \mid x,r,f_n) \cdot Q(dy \mid x,r,f_n), \tag{32}$$

for $n = 0, 1, \dots, N-1$. Then, the deterministic non-stationary policy

$$\pi^* = \{f_0, f_1, \dots, f_{N-1}\}$$

is optimal and the optimal value J^* coincides with J_0, that is,

$$J^*(x,r) = J_0(x,r) = J(\pi^*, x, r), \quad \forall x \in X. \tag{33}$$

Proof. The existence of measurable selectors f_n for each $n = 0, 1, 2, \dots$ is consequence of Theorem 7 in [10] part(ii) with appropriate changes. In this proof, $E^{\pi}_{(x,r)}$ is denoted only by E^{π}, when $(x_0, r_0) = (x,r)$. Let us consider a policy $\pi = \{\pi_n\} \in \Pi^0$, and the functions

$$M_N(\pi,x,r) := E^{\pi}\Big[c(x_N, r_N, a_N) \mid x_N = x, r_N = r \Big] \tag{34}$$

and for $n = 0, 1, \dots, N-1$,

$$M_n(\pi,x,r) := E^{\pi}\Bigg[\sum_{i=n}^{N-1} e^{S_n - S_i} c(x_i, r_i, a_i) + e^{S_n - S_N} c(x_N, r_N, a_N) \mid x_n = x, r_n = r \Bigg]. \tag{35}$$

M_n represents the expected value from step n up to N. As $J(\pi,x,r) < \infty$, then M_n is finite for $n = 0, 1, 2, \dots, N-1$. For $n = N$, from definition (34), we obtain

$$\begin{aligned} M_N(\pi,x,r) &= E^{\pi}[c(x_N, r_N, a_N) \mid x_N = x, r_N = r] \\ &\geq J_N(x,r). \end{aligned}$$

We shall prove that for $n = 0, \dots, N$

$$M_n(\pi, x, r) \geq J_N(x, r),\ \forall (n, r) \in X'.$$

We proceed by backward induction. Suppose the validity of backward induction hypothesis: that is, for all $(x, r) \in X'$

$$M_{n+1}(\pi, x, r) \geq J_{n+1}(x, r). \tag{36}$$

Then,

$$\begin{aligned}
&M_n(\pi, x, r) \\
&= E^\pi \left[\sum_{i=n}^{N-1} e^{S_n - S_i} c(x_i, r_i, a_i) + e^{S_n - S_N} c(x_N, r_N, a_N) \mid x_n = x, r_n = r \right] \\
&= E^\pi \Bigg[c(x_n, r_n, a_n) + \sum_{i=n+1}^{N-1} e^{S_n - S_i} c(x_i, r_i, a_i) \\
&\qquad + e^{S_n - S_N} c(x_N, r_N, a_N) \mid x_n = x, r_n = r \Bigg] \\
&= E^\pi \Bigg[c(x_n, r_n, a_n) \\
&\qquad + E\Bigg[e^{-r_n \delta_{n+1}} \left(\sum_{i=n+1}^{N-1} e^{S_{n+1} - S_i} c(x_i, r_i, a_i) + e^{S_{n+1} - S_N} c(x_N, r_N, a_N) \right) \\
&\qquad\qquad \Big| x_{n+1} = x', r_{n+1} = R_n(r, \delta_{n+1}), a_{n+1} = a' \Bigg] \Bigg| x_n = x, r_n = r \Bigg] \\
&= \int_A \Bigg[c(x, r, a) + \int_X E[e^{-r\delta_{n+1}} M_{n+1}(\pi, x', R_n(r, t)) \mid x, r, a] \\
&\qquad \cdot Q(dx' \mid x, r, a) \Bigg] \pi_n(da \mid x, r) \\
&\geq \int_A \Bigg[c(x, r, a) + \iint\limits_{X\,0}^{\infty} e^{-rt} J_{n+1}(x', R_n(r, t)) G(dt \mid x, r, a) \\
&\qquad \cdot Q(dx' \mid x, r, a) \Bigg] \pi_n(da \mid x, r) \\
&\geq \min_{A(x,r)} \Bigg[c(x, r, a) + \iint\limits_{X\,0}^{\infty} e^{-rt} J_{n+1}(x', R_n(r, t)) G(dt \mid x, r, a) \\
&\qquad \cdot Q(dx' \mid x, r, a) \Bigg] \\
&= J_n(x, r).
\end{aligned}$$

Hence, for all $(x,r) \in X'$,

$$M_n(\pi,x,r) \geq J_n(x,r), \quad n = N-1,\dots,0,$$

and

$$J(\pi,x,r) = M_0(\pi,x,r) \geq J_0(x,r) = J(\pi^*,x,r).$$

Then

$$J^*(x,r) = \inf_{\pi\in\Pi} J(\pi,x,r) \geq J(\pi^*,x,r).$$

The other inequality is follows by the definition of J^*. This proves the Theorem. □

4. THE INFINITE HORIZON SEMI-MARKOV CONTROL PROBLEM

Consider the SMCM (2). The expected discounted total cost with recursive discount rate infinite horizon is defined by

$$V(\pi,x,r) := E^{\pi}_{(x,r)}\left[\sum_{n=0}^{\infty} e^{-S_n} c(x_n,r_n,a_n)\right], \tag{37}$$

for the initial pair $(x_0,r_0) = (x,r)$ and the policy $\pi \in \Pi$, where S_n is given in (27) and c is the total sojourn cost (6).

The optimal value function with infinite horizon is

$$V^*(x,r) := \inf_{\Pi}\ V(\pi,x,r), \quad x \in X. \tag{38}$$

The infinite horizon *semi-Markov Control problem* is to find a policy π^* such that

$$V(\pi^*,x,r) = V^*(x,r),$$

holds for all $x \in X$. In such a case, π^* is an optimal policy. The existence and measurability of V^* are obtained by convergence of semi-Markov control problems with finite horizon [9]. Thus, the conditions used in the finite horizon case are also assumed here. We introduce the following additional notation.

Definition 4.1. Let m and n be the non-negative numbers, such that $m \leq n$. For the discount rate $r = r_m$, *the expected total cost from step m up to step n* is defined by

$$V^n_m(\pi,x,r) = E^{\pi}_{(x_0,r_0)}\left[\sum_{t=m}^{n} e^{-(S_t-S_m)} c(x_t,a_t) \mid x_m = x, r_m = r\right], \tag{39}$$

for any $\pi \in \Pi$ and $x \in X$. The value function from step m up to step n is

$$V^{n*}_m(x,r) := \inf_{\Pi}\ V^n_m(\pi,x,r), \quad \forall x \in X,\ r = r_m. \tag{40}$$

The semi-Markov Control problem from step m up to step n is to find a policy $\pi^{(m,n)*} \in \Pi$ such that

$$V^{n*}_m(\pi^{(m,n)*},x,r) = V^{n*}_m(x,r), \quad \forall x \in X,\ r = r_m. \tag{41}$$

For $m = 0$, $V^n_0(\pi,x,r) := V^n(\pi,x,r)$. Let m be fixed,

$$V_m(\pi,x,r) := \lim_{n\to\infty} V^n_m(\pi,x,r).$$

Assumption 4.2.

1. There exists a policy $\pi \in \Pi$ such that $V(\pi, x, r) < \infty$ for any $(x, r) \in X'$. The family of policies that satisfies this condition is denoted by Π^1.

Definition 4.3. Let $\epsilon > 0$. A policy $\pi_\epsilon \in \Pi^1$ is $\epsilon-$*optimal* if

$$V(\pi_\epsilon, x, r) \leq \inf_{\pi \in \Pi^1} V(\pi, x, r) + \epsilon, \ \forall (x, r) \in X'.$$

Lemma 4.4. Let $x \in X$, $r = r_m$ and $m = 0, 1, 2, \dots$ Under Assumptions 3.2 and 4.2,

$$\lim_{n \to \infty} V_m^{n*}(x, r) = V_m^*(x, r). \tag{42}$$

Proof. Let us define, for each $k = 0, 1, 2, \dots$, the measurable functions on X'

$$u^0(x, r_k) := \inf_{a \in A(x, r_k)} c(x, r_k, a), \tag{43}$$

and for $n = 1, 2, \dots$

$$u^n(x, r_k) := \inf_{a \in A(x, r_k)} \Big[c(x, r_k, a) + \iint\limits_{X\,0}^{\quad\infty} e^{-r_k t_{k+1}} u^{(n-1)}(y, R_k(r_k, t_{k+1})) \tag{44}$$
$$\cdot\, G(dt_{k+1} \mid x, r_k, a) Q(dy \mid x, r_k, a) \Big].$$

We shall show that $u^n(x, r_k) = V_k^{n*}(x, r_k)$ for all $n = 0, 1, 2, \dots$, $k = 0, 1, 2, \dots$, and $x \in X$. It is proved by induction over n.

For $n = 0$, note that

$$u^0(x, r_k) \leq V_k^k(\pi, x, r_k), \ \ \forall \pi \in \Pi, \ x \in X, \ k = 0, 1, 2, \dots$$

Taking the infimum over Π,

$$u^0(x, r_k) \leq V_k^{k*}(x, r_k), \ k = 0, 1, 2, \dots \tag{45}$$

On the other hand, by the measurable selector condition (see Theorem 3.4), there exists for each $k = 0, 1, 2, \dots$, a selector f_k^0 such that

$$\begin{aligned} u^0(x, r_k) &= c(x, r_k, f_k^0(x, r_k)) \\ &= V_k^k(\pi^{k,k}, x, r_k) \\ &\geq V_k^{k*}(x, r_k), \end{aligned} \tag{46}$$

where $\pi^{k,k}$ represents the policy $\pi^{k,k} = \{\pi_0, \dots, \pi_{k-1}, f_k^0, \pi_{k+1}, \dots\}$. Then, by the inequalities (45) and (46) we conclude that $V_k^{k*} = u^0$ for every $x \in X$ and $k = 0, 1, 2, \dots$ Let us now assume for each $k = 0, 1, 2, \dots$, that the induction hypothesis holds for n, i.e.,

$$u^n(x, r_k) = V_k^{(n+1)*}(x, r_k), \ k = 0, 1, 2, \dots$$

Then, for each $k = 0, 1, 2, \ldots$, there exists a measurable selector f_k^n such that

$$u^n(x, r_k) = c(x, r_k, f_k^n) + \int_X \int_0^\infty e^{-r_k t_{k+1}} u^{(n-1)}(y, R_k(r_k, t_{k+1})) \\ \cdot G(dt_{k+1} \mid x, r_k, f_k^n) Q(dy \mid x, r_k, a),$$

$k = 0, 1, 2, \ldots$ and the policy

$$\pi^{k,n} = \{\pi_0, \ldots, \pi_{k-1}, f_k^n, \ldots, f_{k+n}^0, \pi_{k+n+1}, \ldots\}$$

satisfies

$$V_k^n(\pi^{k,n}, x, r_k) = V_k^{n*}(x, r_k),\ k = 0, 1, 2, \ldots$$

For $n + 1$, we obtain

$$u^{n+1}(x, r_k) \leq V_k^{n+1}(\pi, x, r_k), \quad \forall \pi \in \Pi,\ x \in X,\ k = 0, 1, 2, \ldots$$

and taking the infimum over Π,

$$u^{n+1}(x, r_k) \leq V_k^{(n+1)*}(x, r_k),\ k = 0, 1, 2, \ldots \tag{47}$$

Again, by the measurable selector condition there exists, for each $k = 0, 1, 2, \ldots$, a selector f_k^{n+1} such that

$$\begin{aligned} u^{n+1}(x, r_k) &= c(x, r_k, f_k^{n+1}) + \int_X \int_0^\infty e^{-r_k t_{k+1}} u^{n-1}(y, R_k(r_k, t_{k+1})) \\ &\qquad \cdot G(dt_{k+1} \mid x, r_k, f_k^{n+1}) Q(dy \mid x, r_k, a), \\ &= V_k^{n+1}(\pi^{k,n+1}, x, r_k) \\ &\geq V_k^{(n+1)*}(x, r_k), \end{aligned}$$

where $\pi^{k,n+1}$ is the policy

$$\pi^{k,n+1} = \{\pi_0, \ldots, \pi_{k-1}, f_k^{n+1}, f_{k+1}^n, \ldots, f_{k+n}^1, f_{k+n+1}^0, \pi_{k+2}, \ldots\}.$$

Hence, $V_k^{n+1*}(x, r_k) = u^{n+1}(x, r_k)$ for any $x \in X$, $k = 0, 1, 2, \ldots$ and $n = 0, 1, 2, \ldots$

By construction of u^n

$$0 \leq u^n(x, r_k) = V_k^{n*}(x, r_k) \leq u^{n+1}(x, r_k) \leq V_k^*(x, r_k), \quad \text{for } k, n = 0, 1, 2, \ldots$$

Then there exists a measurable function $\overline{U}_k$ over X' such that $\overline{U}_k(x, r_k) \leq V_k^*(x, r_k)$, $k = 0, 1, 2, \ldots$ and $u^n \uparrow \overline{U}_k$. Note by the convergence of u^n to $\overline{U}_k$, $u^n = V_k^{(n)*}$ for all k, all n and $V_k^* \downarrow 0$, $k \to \infty$. Additionally, for all k is valid for the sequence of selectors $f^{\infty,0} = \{f_k, f_{k+1}, \ldots\}$ in the policy $\pi^{k,\infty} := \{\pi_0, \ldots, \pi_{k-1}, f_k, f_{k+1}, \ldots\} \in \Pi^1$ that the inequality

$$\overline{U}_k(x, r_k) \geq V_k(\pi^{k,\infty}, x, r_k),$$

is valid for all $x \in X$, and by definition of V_k^*

$$V_k^*(x, r_k) = V_k(\pi^{0,\infty}, x, r_k).$$

i. e., $V_k^*(x, r_k) \leq \overline{U}_k(x, r_k)$ for all $x \in X'$. Therefore, $\overline{U}_k(x, r_k) = V_k^*(x, r_k)$, for all $k = 0, 1, 2, \ldots$ and all $x \in X'$. □

Theorem 4.5. Under Assumption 4.2, there exists a deterministic non-stationary policy $\pi \in \Pi^1$, such that,

$$V(\pi, x, r) = V^*(x, r),\ \forall (x, r) \in X'.$$

Proof. Notice that, by the Lemma 4.4, the functions V_n^* is measurable, $n \in \mathbb{N}$.

We shall prove that

$$V_n^*(x, r_n) = \inf_{a \in A(x, r_n)} \Big[c(x, r_n, a) + \iint\limits_{X\,0}^{\quad\infty} e^{-r_n t_{n+1}} V_{n+1}^*(y, R_n(r_n, t_{n+1})) \cdot G(dt_{n+1} \mid x_n, r_n, a) Q(dy \mid x, r_n, a) \Big], \tag{48}$$

for all $n = 0, 1, 2, \ldots$ Let us define

$$\nu(x, r_n) := \inf_{a \in A(x, r_n)} \Big[c(x, r_n, a) + \iint\limits_{X\,0}^{\quad\infty} e^{-r_n t_{n+1}} V_{n+1}^*(y, R_n(r_n, t_{n+1})) \cdot G(dt_{n+1} \mid x_n, r_n, a) Q(dy \mid x, r_n, a) \Big],$$

$n = 0, 1, 2, \ldots$ Thus, for any policy $\pi \in \Pi$

$$V_n(\pi, x, r_n) \geq V_n^*(x, r_n),\quad V_n(\pi, x, r_n) \geq \nu(x, r_n),\ \forall n,$$

and therefore

$$\begin{aligned} &V_n(\pi, x, r_n) \\ &= E_{(x, r_0)}^{\pi} \Big[c(x, r_n, a) + \iint\limits_{X\,0}^{\quad\infty} e^{-r_n t_{n+1}} V_{n+1}(y, r_{n+1}) \times G(dt_{n+1} \mid x_n, r_n, a) \cdot Q(dy \mid x, r_n, a) \Big] \\ &\geq E_{(x, r_n)}^{\pi} \Big[c(x, r_n, a) + \iint\limits_{X\,0}^{\quad\infty} e^{-r_n t_{n+1}} V_{n+1}^*(y, r_{n+1}) G(dt_{n+1} \mid x_n, r_n, a) \cdot Q(dy \mid x, r_n, a) \Big] \end{aligned}$$

$$\geq \inf_{a\in A(x,r_n)} \Big[c(x,r_n,a) + \iint\limits_{X\,0}^{\infty} e^{-r_n t_{n+1}} V_{n+1}^*(y, R_n(r_n,t_{n+1})) \\ \cdot G(dt_{n+1} \mid x_n, r_n, a) Q(dy \mid x, r_n, a) \Big],$$

and taking the infimum over Π,

$$V_n^*(x,r_n) \geq \nu(x,r_n). \tag{49}$$

On the other hand, if $\pi \in \Pi^1$

$$V_n^*(x,r_n) \leq V_n(\pi,x,r_n),$$

then $V_n(\pi,x,r_n) \downarrow 0$, $n\to\infty$. Consider now the policy

$$\pi^{n,\infty} = \{\pi_0,\dots,\pi_{n-1}, f_n, f_{n+1},\dots\},$$

where $(f_i,\ i=n,n+1,\dots)$ is a measurable selector, which exists by Theorem 7 in [10]. By Lemma 4.4, for any $\epsilon > 0$, there exists m_0 such that, for any $m > m_0$ and each $n = 0,1,2,\dots$

$$\begin{aligned} V_n(\pi^{n,\infty},x,r_n) &< V_n^m(\pi^{n,\infty},x,r_n) + \epsilon \\ &= V_n^{m*}(x,r_n) + \epsilon, \end{aligned}$$

and hence $V_n(\pi^{n,\infty},x,r_n) \leq V_n^*(x,r_n)$, $m\to\infty$.
On the other hand, by definition of V_n^*, $V_n(\pi^{n,\infty},x,r_n) \geq V_n^*(x,r_n)$, and

$$V_n(\pi^{n,\infty},x,r_n) = V_n^*(x,r_n),\ \forall n = 0,1,2,\dots$$

Moreover

$$\begin{aligned} \nu(x,r_n) &= c(x,r_n,f_n) + \iint\limits_{X\,0}^{\infty} e^{-r_n t_{n+1}} V_{n+1}^*(y, R_n(r_n,t_{n+1})) \\ &\qquad \cdot G(dt_{n+1} \mid x_n, r_n, f_n) Q(dy \mid x, r_n, f_n) \\ &= c(x,r_n,f_n) + \iint\limits_{X\,0}^{\infty} e^{-r_n t_{n+1}} V_{n+1}(\pi^{n,\infty}, x_{n+1}, r_{n+1}) \\ &\qquad \cdot G(dt_{n+1} \mid x_n, r_n, f_n) Q(dy \mid x, r_n, f_n) \\ &\geq V_n^*(x,r_n), \end{aligned}$$

and by inequality (49), the equality (48) holds. Hence, the deterministic non-stationary policy

$$\pi^{0,\infty} := \{f_0, f_1, f_2, \dots\}$$

satisfies, for $n = 0$ in (48), that

$$V^*(x,r) = V_0^*(x,r_0)$$

$$
\begin{aligned}
&= c(x, r_0, f_0) + \iint\limits_{X\,0}^{\infty} e^{-r_0 t_1} V_1^*(x_1, r_1) \times G(dt_{n+1} \mid x_n, r_n, f_0) \\
&\qquad \cdot Q(dy \mid x, r_n, f_0) \\
&= V(\pi^{f^{0,\infty}}, x, r_0)
\end{aligned}
$$

Therefore, $\pi^{f^{0,\infty}}$ is an optimal policy. □

Theorem 4.6. Suppose Assumption 4.2 holds. If there exists a policy $\overline{\pi}$ such that

$$
\begin{aligned}
V_n(\overline{\pi}, x, r_n) = \inf_{a \in A(x, r_n)} \Big[c(x,r_n, a) + \iint\limits_{X\,0}^{\infty} e^{-r_n t_{n+1}} V_{n+1}(\overline{\pi}, y, R_n(r_n, t_{n+1})) \\
\cdot G(dt_{n+1} \mid x_n, r_n, a) Q(dy \mid x, r_n, a) \Big],
\end{aligned} \tag{50}
$$

for all $x \in X,\ n = 0, 1, 2, \ldots$, and satisfies

$$
\lim_{n \to \infty} e^{-S_n} E^{\pi}_{(x,r)} V_n(\overline{\pi}, x_n, r_n) = 0, \quad (x, r) \in X', \pi \in \Pi^1, \tag{51}
$$

then, $V(\pi, x, r) = V^*(x, r)$ for all $(x, r) \in X'$.

Proof. If (50) holds, then by definition of V^*, $V(\pi, x, r) \geq V^*(x, r)$ for all $(x, r) \in X'$. On the other hand, for any $\pi \in \Pi$, and $(x, r) \in X'$, the Markov Property yields,

$$
\begin{aligned}
&E^{\pi}_{(x,r)}[e^{-S_{n+1}} V_{n+1}(\overline{\pi}, x_{n+1}, r_{n+1}) \mid h_n, a_n] \\
&= E^{\pi}_{(x,r)} \Big[e^{-S_n} \big[c(x_n, r_n, a_n) + \iint\limits_{X\,0}^{\infty} e^{-r_n t_{n+1}} V_{n+1}(\overline{\pi}, y, r_{n+1}) \\
&\qquad \cdot G(dt_{n+1} \mid x_n, r_n, a) Q(dy \mid x, r_n, a) - c(x_n, r_n, a_n) \big] \Big] \\
&\geq E^{\pi}_{(x,r)} \Big[e^{-S_n} \Big[V_n(\overline{\pi}, x_n, r_n) - c(x_n, r_n, a_n) \Big] \Big],
\end{aligned}
$$

and so

$$
\begin{aligned}
&E^{\pi}_{(x,r)}[e^{-S_n} c(x_n, r_n, a_n)] \\
&\geq -E^{\pi}_{(x,r)} \left[e^{-S_{n+1}} [V_{n+1}(\overline{\pi}, x_{n+1}, r_{n+1}) - e^{-S_n} V_n(\overline{\pi}, x_n, r_n) \mid h_n, a_n) \right].
\end{aligned}
$$

Thus, taking expectations $E^{\pi}_{(x,r)}$ and adding over $i = 0, 1, \ldots, k-1$,

$$
E^{\pi}_{(x,r)} \sum_{i=0}^{k-1} e^{-S_i} c(x_i, r_i, a_i) \geq V_0(\overline{\pi}, x, r) - e^{-S_k} E^{\pi}_{(x,r)} V_k(\overline{\pi}, x_k, r_k),
$$

holds for every k. Letting $k \to \infty$ and using (51), it follows that $V(\pi, x, r) \geq V(\pi, x, r)$ for all $(x, r) \in X'$, and $V^*(x, r) \geq V(\pi, x, r)$ for every $(x, r) \in X'$. In consequence, $V^*(x, r) = V(\pi, x, r)$ for all $(x, r) \in X'$. □

Proposition 4.7. Under Assumption 4.2, there exist deterministic non-stationary $\epsilon-$optimal policies to the semi-Markov control problem with infinite horizon.

Proof. Let $\tilde{\pi} \in \Pi^1$ be a fixed policy, $\epsilon > 0$. For any $(x, r) \in X'$, consider the measurable functions

$$u^0(x,r) = \inf_{a \in A(x,r)} c(x,r,a) \tag{52}$$

and, $n = 1, 2, \ldots,$

$$u^n(x,r) = \inf_{a \in A(x,r_n)} \Bigg[c(x,r_n,a) + \iint\limits_{X\,0}^{\quad\infty} e^{-r_n t_{n+1}} u^{n-1}(y, r_{n+1}) \tag{53}$$

$$\cdot\, G(dt_{n+1} \mid x_n, r_n, a) Q(dy \mid x, r_n, a) \Bigg].$$

By Condition (4.2), for any $n \in \mathbb{N}$, exist the sequence of selectors $\{f_0, \ldots, f_n\}$ such that the policy $\pi^{(n)} := \{f_0, \ldots, f_n, \tilde{\pi}_{n+1}, \tilde{\pi}_{n+2}, \ldots\}$ satisfies

$$u^n(x,r) \le V^n(\pi^{(n)}, x, r) \le V(\tilde{\pi}, x, r).$$

Then, for some $n^* \in \mathbb{N}$,

$$V(\pi^{(n^*)}, x, r) \le V(\pi, x, r) + \epsilon, \ \forall \pi \in \Pi^1, \ \forall (x,r) \in X'.$$

Taking infimum over π,

$$V(\pi^{(n)}, x, r) \le \inf_{\pi \in \Pi^1} V(\pi, x, r) + \epsilon, \ \forall (x,r) \in X'.$$

Therefore, for $n > n^*$, $\pi^{(n)}$ are deterministic non-stationary $\epsilon-$optimal policies. □

Remark 4.8.

1. By Lemma 4.4, the function V^{n*} is measurable for each n and then the $\lim_{n\to\infty} V^{n*}$ is also measurable.

2. Observe that for $\pi \in \Pi^1$ and any natural number n

$$\lim_{k\to\infty} \left[\exp\left\{ -\sum_{i=n}^{k} r_i \delta_{i+1} \right\} V^{n+k}(\pi, x_{n+k}, r_{n+k}) \right] = 0. \tag{54}$$

5. AN EXAMPLE OF A SMCM WITH DISCOUNT RATE RECURSIVE FUNCTIONS

One way of getting recursive discount rate functions $\{R_n \mid n = 0, 1, 2, \dots\}$ is using continuous time stochastic processes $\{\hat{r}_t \mid n \le t \le n+1, \}$, $n = 0, 1, 2, \dots$, to represent the discount rate between consecutive steps of the SMCM. In this case, it is supposed that the dynamics follows the scalar stochastic differential equation (SDE) given by

$$d\hat{r}_t = \kappa_n\big(\theta_n - \hat{r}_t\big)dt + \sigma_n dW_t, \quad t \in [n, n+1], \quad n = 0, 1, 2, \dots, \tag{55}$$

where $\kappa_n \in [\kappa_{min}, \kappa_{max}]$, $\sigma_n \in [\sigma_{min}, \sigma_{max}]$, $\theta_n \in [\theta_{min}, \theta_{max}]$ are positive constants, θ_n is the mean value or tendency of the SDE, and the initial condition is the previous discount rate $\hat{r}_n$ (if $t \in [0, 1]$, the initial condition is the discount rate r_0). Similar to some classical short-rate models (see [21]), the SDE (55) has the mean-reverting property: the discount rate $\hat{r}_t$ tends to fluctuate around of ideal value θ_n when $t \in [n, n+1]$, $n = 0, 1, 2, \dots$ To guarantee the existence of a unique solution to the SDE, we suppose that the functions $\mu_n(t, z) := \kappa_n[\theta_n - z]$ and $\sigma_n(t, z) = \sigma_n$, $n \ge 0$ satisfy the classical *Itô* conditions [1, Chapter 6], i. e., exists $K > 0$, such that

1. For all $t \in [n, n+1]$, $n = 0, 1, 2, \dots$, $z_1, z_2 \in \mathbb{R}$,

$$|\mu(t, z_1) - \mu(t, z_2)| + |\sigma(t, z_1) - \sigma(t, z_2)| \le K|z_1 - z_2|. \tag{56}$$

2. For all $t \in [n, n+1]$, $n = 0, 1, 2, \dots$, and $z \in \mathbb{R}$

$$|\mu(t, z)|^2 + |\sigma(t, z)|^2 \le K^2\big(1 + |z|^2\big). \tag{57}$$

The solution is the mean-reverting stochastic processes

$$\hat{r}_t = r_n e^{-(t-n)\kappa_n} + \theta_n[1 - e^{-(t-n)\kappa_n}] + \sigma_n e^{-(t-n)\kappa_n} \int\limits_n^t e^{u\kappa_n} dW_u, \tag{58}$$

where $t \in [n, n+1]$, W_u, $u \in [n, t]$ is the Wiener process, θ_n is the mean reverting value of processes and the κ_n is this the velocity of reversion.

From the solution (58), by taking the expectation and substituting the time-dependent expression by the sojourn time variable δ_{n+1}, we obtain the recursive equation

$$\begin{aligned} r_{n+1} &:= E[R_n(r_n, \delta_{n+1}) \mid x_n, r_n, a_n] \\ &= \int\limits_0^\infty \big(r_n \exp\big(-t\kappa_n\big) + \theta_n\big[1 - \exp\big(-t\kappa_n\big)\big]\big)\, G(dt \mid x_n, r_n, a_n), \end{aligned} \tag{59}$$

where $G(\cdot \mid x_n, r_n, a_n)$ is the sojourn time distribution of δ_{n+1} conditioned by (x_n, r_n, a_n), for each $n = 0, 1, 2, \dots$, and r_0 is the initial discount rate.

6. LINEAR-QUADRATIC SEMI-MARKOV MODEL.

Let us consider an SMCM with $X = A = \mathbb{R}$, where the evolution of states is given by

$$x_{n+1} = Ax_n + Ba_n + \xi_n, \quad n = 0, 1, 2, \ldots,$$

A and B are non-negative constants and the random variables ξ_n are supposed i.i.d., independent of x_0 such that $E\xi_n = 0$ and $0 < E\xi_n^2 = \sigma^2 < \infty$. The sojourn times δ_n, $n = 1, 2, \ldots$, are random variables on $(0, \infty)$, independent of x, with exponential distribution function G on $\mathbb{K} = \mathbb{R} \times (0, \infty) \times \mathbb{R}$, where its parameter is given by the non-negative continuous function $g : \mathbb{K} \to \mathbb{R}$. The immediate and sojourn costs (see (19)) are given by

$$\begin{aligned} D(x,a) &= q_2 a, \\ d(x,a) &= \alpha_1 x + \alpha_2 a^2, \\ c(x,r,a) &= D(x,a) + \tau(x,r,a) d(x,a) \\ &= D(x,a) + \frac{1}{r}\left(1 - \frac{g(x,r,a)}{r(1+g(x,r,a))}\right) d(x,a), \end{aligned}$$

where $q_2, \alpha_1, \alpha_2 > 0$, and $(x_n, r_n, a_n) \in \mathbb{K},\ n = 0, 1, 2, \ldots$

The evolution of the discount rates follows the SDE (55) and satisfies the recursive relation (59). The discounted total cost in this case takes the form (37).

The semi-Markov linear quadratic control problem with finite horizon consists of finding an optimal policy in N steps such that

$$J(\pi, x, r) = E^{\pi}_{(x,r)}\left[\sum_{n=0}^{N-1} e^{-S_n} c(x_n, r_n, a_n)\right]$$

is minimized, where $S_0 = 0$ and S_n as (27). The dynamic programming equations are

$$J_N(x_N, r_N) = \min_{a \in A(x_N, r_N)} c(x_N, r_N, a) \tag{60}$$

and for $n = N-1, N-2, \ldots, 0$

$$J_n(x, r_n) = \min_{a \in A(x_n, r_n)} \Bigg[c(x_n, r_n, a) + E\Big[e^{-r_n \delta_{n+1}} J_{n+1}(Ax_n + Ba + \xi_n, R_n(r_n, \delta_{n+1})) \mid x_n = x, r_n, a_n = a \Big] \Bigg] \tag{61}$$

which become

$$J_n(x, r_n) = \min_{a \in A(x_n, r_n)} \Bigg[q_2 a + \frac{\alpha_1 x_n^2 + \alpha_2 a^2}{r_n + g(x_n, r_n, a)} + \int_0^{\infty} E\Big[g(x_n, r_n, a) e^{-s(r_n + g(x_n, r_n, a))} \cdot J_{n+1}(Ax_n + Ba + \xi_n, (r_n - \theta_n) e^{-s\kappa_n} + \theta_n) \Big] ds \Bigg]. \tag{62}$$

By Theorem 3.4, exists a non-stationary optimal policy $\pi^{N*} = \{f_i\}_{i=0}^{N}$ and the optimal value is $J_0(x_0, r_0)$.

An example with explicit solution, can be obtained when the recursive discount rate is

$$r_{n+1} = E[R_n(r_n, \delta_{n+1}) \mid x_n, r_n, a_n] := r_n e^{-r_n k \delta n+1}, \ k > 0,$$

the function $g(x, r, a) = \frac{r\lambda}{r+(1-r)(\lambda+1)}$, and the sojourn time distribution is reduced to

$$G(t \mid x, r, a) := \int\limits_0^t r_n \lambda e^{-r_n \lambda s} ds,$$

$\lambda > 0$, for all $(x, r, a) \in \mathbb{K}$. In this case, from the dynamic programming equations (60) and (61) , if $n = N$, the minimum of J_N^* occurs when $a_N = -r_N \frac{q_2(\lambda+1)}{2\alpha_2}$ and

$$J_N(x, r_N) = \frac{1}{r_N} I_N x - r_N H_N,$$

where $H_N := \frac{q_2^2(\lambda+1)}{4\alpha_2}$ and $I_N := \frac{\alpha_1}{\lambda+1}$. For $n = N-1, N-2, N-3, \ldots, 0$ we obtain

$$\begin{aligned} J_n(x, r_n) = \min_{a \in A(x, r_n)} \Bigg[& \frac{1}{r_n} \cdot \frac{\alpha_2}{\lambda+1} a^2 + \left(q_2 + \frac{I_{n+1}\lambda B}{r_n(\lambda+1-k)} \right) a + \frac{1}{r_n} I_n x \\ & - r_n \Big(H_n \frac{\lambda}{\lambda+1+k} \Big) + \frac{1}{r_n} \cdot \frac{T_n \lambda}{\lambda+1-k} - \frac{U_n \lambda}{\lambda+1} \Bigg], \end{aligned}$$

where $I_n := \frac{I_{n+1}\lambda A}{\lambda+1-k} + \frac{\alpha_1}{\lambda+1}$, $H_n = H_N + H_{n+1}\frac{\lambda}{\lambda+1+k}$, $T_n := \frac{(I_{n+1}\lambda B)^2(1+\lambda)}{4\alpha_2(\lambda+1-k)^2} + T_{n+1}\frac{\lambda}{\lambda+1-k}$, $U_n := \frac{q_2 I_{n+1}\lambda(\lambda+1)B}{2\alpha_2(\lambda+1-k)} + U_{n+1}\frac{\lambda}{\lambda+1}$, and $T_N = U_N := 0$. By standard calculus arguments, the minimum is

$$a_n = -\frac{q_2 + I_{n+1}\lambda B}{\frac{2\alpha_2}{r_n(\lambda+1)}} := f_n(x, r_n). \tag{63}$$

and the optimal value is

$$J_n^*(x, r_n) = \frac{1}{r_n} I_n x - r_n H_n - \frac{1}{r_n} T_n - U_n.$$

This iterative process generates the non-stationary policy

$$f^* = \{a_0, a_1, \ldots, a_N\}$$

and the value function $J^*(x, r) = J_0(x, r)$, for all $(x, r) \in X'$.

The expected discounted total cost with infinite horizon in this case is

$$V(\pi, x, r) = E_{(x,r)}^{\pi} \left[\sum_{n=0}^{\infty} e^{-S_n'} c(x_n, r_n, a_n) \right].$$

Let $\epsilon > 0$. To obtain a non-stationary $\epsilon-$optimal policy, we consider the non-stationary policy $\pi' = \{f'_0, f'_1, f'_2, \dots\}$ given by measurable selectors

$$f'_i(x,r) = \frac{q_2 d_2(1+\lambda)}{2\lambda_2}, \ i = 0, 1, 2, \dots$$

and is valid that $\pi' \in \Pi^1$:

$$\begin{aligned} V(\pi', x, r) &= E^{\pi'}_{(x,r)}\left[\sum_{n=0}^{\infty} e^{-S'_n} c(x_n, r_n, a_n)\right] \\ &< \Big(2\frac{q_2^2 d_2^2(1+\lambda)^2}{4\alpha_2^2} + \frac{\alpha_1}{\alpha_2}x\Big) E^{\pi'}_{(x,r)}\left[\sum_{n=0}^{\infty} e^{-S'_n}\right] < \infty. \end{aligned}$$

Let $\pi^{n*} = \{a_0, a_1, a_2, \dots, a_n\}$ be a non-stationary optimal policy of the finite horizon case. Hence, for the policy

$$\hat{\pi} = \{a_0, a_1, \dots, a_n, f'_{n+1}, f'_{n+2}, f'_{n+3}, \dots\},$$

there exists a natural number n_0 such that if $n > n_0$

$$V^{(n)}(\hat{\pi}, x, r) - V^{(n+1)}(\hat{\pi}, x, r) < \epsilon/2^n,$$

i. e., $V^n(\hat{\pi}, x, r)$ approximates to $V^*(x, r)$ and $\hat{\pi}$ is a non-stationary policy which is $\epsilon-$optimal.

(Received August 3, 2018)

REFERENCES

[1] L. Arnold: Stochastic Differential Equations. John Wiley and Sons, New York 1973.

[2] R. Ash and C. Doléans-Dade: Probability and Measure Theory. Academic Press, San Diego, 2000.

[3] R. Bhattacharya and M. Majumdar: Controlled semi-Markov models - the discounted case. J. Statist. Plann. Inference *21* (1989), 3, 365–381. DOI:10.1016/0378-3758(89)90053-0

[4] D. Bertsekas and S. Shreve: Stochastic Optimal Control: The Discrete Time Case. Athena Scientific, Belmont, Massachusetts 1996.

[5] D. Blackwell: Discounted dynamic programming. Ann. Math. Statist. *36*, (1965), 226–235. DOI:10.1214/aoms/1177700285

[6] J. De Cani: A dynamic programming algorithm for embedded Markov chains the planning horizon is infinitely. Management. Sci. *10* (1963), 716–733. DOI:10.1287/mnsc.10.4.716

[7] R. Drenyovszki, L. Kovács, K. Tornai, A. Oláh, and I. Pintér I.: Bottom-up modeling of domestic appliances with Markov chains and semi-Markov processes. Kybernetika *53* (2017), 6, 1100–1117. DOI:10.14736/kyb-2017-6-1100

[8] R. Dekker and A. Hordijk: Denumerable semi-Markov decision chains with small interest rates. Ann. Oper. Res. *28* (1991), 185–212. DOI:10.1007/bf02055581

[9] Y. García and J. González-Hernández: Discrete-time Markov control process with recursive discounted rates. Kybernetika *52* (2016), 403–426. DOI:10.14736/kyb-2016-3-0403

[10] J. González-Hernández, R. López-Martínez, and J. Pérez-Hernández: Markov control processes with randomized discounted cost. Math. Meth. Oper. Res. *65* (2006), 27–44. DOI:10.1007/s00186-006-0092-2

[11] J. González-Hernández and C. Villarreal-Rodríguez: Optimal solutions of constrained discounted semi-Markov control problems. Bol. Soc. Mat. Mexicana *1* (2013), 19, 111–124.

[12] O. Hernández-Lerma and J. Lasserre: Discrete-Time Markov Control Processes. Basic Optimality Criteria. Springer-Verlag, New York 1996. DOI:10.1007/978-1-4612-0729-0_1

[13] Q. Hu and W. Yue: Markov Decision Processes With Their Applications. Springer-Verlag, Advances in Mechanics and Mathematics book series *14*, (2008). DOI:10.14736/kyb-2017-1-0059

[14] X. Huang and Y. Huang: Mean-variance optimality for semi-Markov decision processes under first passage criteria. Kybernetika *53* (2017), 1, 59–81. DOI:10.14736/kyb-2017-1-0059

[15] R. Howard: Semi-Markovian decision processes. Bull. Int. Statist. Inst. *40* (1963), 2, 625–652.

[16] W. Jewell: Markov-renewal programming I: formulation, finite return models, Markov-renewal programming II: infinite return models, example. Oper. Res. *11* (1963), 938–971. DOI:10.1287/opre.11.6.938

[17] F. Luque-Vázquez and O. Hernández-Lerma: Semi-Markov control models with average costs. Appl. Math. *26* (1999), 315–331. DOI:10.4064/am-26-3-315-331

[18] F. Luque-Vásquez and J. A. Minjárez-Sosa: Semi-Markov control processes with unknown holding times distribution under a discounted criterion. Math. Methods Oper. Res. *61* (2005), 455–468. DOI:10.1007/s001860400406

[19] F. Luque-Vásquez, J. Minjárez-Sosa, and L. Rosas: Semi-Markov control processes with unknown holding times distribution under an average cost criterion. Appl. Math. Optim. *61*, (2010), 317–336. 10.1007/s00245-009-9086-9

[20] P. Schweitzer: Perturbation Theory and Markovian Decision Processes. Ph.D. Dissertation, Massachusetts Institute of Technology, 1965.

[21] O. Vasicek: An equilibrium characterization of the term structure. J. Financ. Econom. *5* (1977), 177–188. DOI:10.1016/0304-405x(77)90016-2

[22] O. Vega-Amaya: Average optimatily in semi-Markov control models on Borel spaces: unbounded costs and control. Bol. Soc. Mat. Mexicana *38* (1997), 2, 47–60.

[23] R. Zagst: The effect of information in separable Bayesian semi-Markov control models and its application to investment planning. ZOR – Math. Methods Oper. Res. *41* (1995), 277–288.

Yofre H. García, Ciudad Universitaria, Carretera Emiliano Zapata km. 8, Rancho San Francisco, Facultad en Física y Matemáticas, Universidad Autónoma de Chiapas, Terán, C. P. 29050, Tuxtla Gutiérrez, Chiapas. México.
e-mail: yofregarcia@gmail.com

Juan González-Hernández, Avenida Universidad 3000, Ciudad Universitaria, Circuito Escolar S/N, IIMAS-UNAM, Departamento de Probabilidad y Estadística, Coyoacán, C.P. 04510, Distrito Federal. México.
e-mail: juan@sigma.iimas.unam.mx