Acta Appl Math (2011) 114: 135–156
DOI 10.1007/s10440-011-9605-y

Semi-Markov Control Models with Partially Known Holding Times Distribution: Discounted and Average Criteria

Fernando Luque-Vásquez · J. Adolfo Minjárez-Sosa · Luz del Carmen Rosas-Rosas

Received: 19 February 2010 / Accepted: 27 February 2011 / Published online: 30 March 2011

Abstract The paper deals with a class of semi-Markov control models with Borel state and control spaces and possibly unbounded costs, where the holding times distribution F depends on an unknown and possibly non-observable parameter which may change from stage to stage. The system is modeled as a game against nature, which is a particular case of a minimax control system. The objective is to show the existence of minimax strategies under the discounted and average cost criteria.

Keywords Semi-Markov control processes · Discounted and average cost criteria · Minimax control systems · Games against nature

Mathematics Subject Classification (2000) 90C40 · 90C47 · 93E20

1 Introduction

The paper deals with semi-Markov control models (SMCMs) with Borel state and control spaces, possibly unbounded costs, and distribution of the holding times partially known by the controller, under discounted and average cost criteria.

The SMCMs are a class of continuous time stochastic control models where the distribution of the random times between consecutive decision epochs (holding or sojourn times) is arbitrary. Generally, such distribution depends on the state and the control selected by the controller in each decision epoch, but in this paper we suppose that additionally it also depends on an unknown and possibly non-observable parameter which may change from

F. Luque-Vásquez · J.A. Minjárez-Sosa (✉) · L.C. Rosas-Rosas
Departamento de Matemáticas, Universidad de Sonora, Rosales s/n, Col. Centro, 83000 Hermosillo, Sonora, Mexico
e-mail: aminjare@gauss.mat.uson.mx

F. Luque-Vásquez
e-mail: fluque@gauss.mat.uson.mx

L.C. Rosas-Rosas
e-mail: lcrosas@gauss.mat.uson.mx

stage to stage. In this sense, the distribution of the holding times is partially known by the controller.

In previous works the authors have studied a similar problem in which is assumed that the holding time distribution is unknown, but with a density independent of the state-action pairs (see [18, 19]). Under this assumption it was possible to apply standard schemes consisting in the combination of statistical density estimation methods with control procedures to construct optimal strategies. In general, when it is possible, the unknown component is estimated from historical data using statistical methods. Then the optimal strategy is computed assuming that such estimate is the true component, and then the well-known *principle of estimation and control*, proposed in [15] and [20], is applied. However, in our model, because the dependence on the state-action pairs, as well as the non-observability of the unknown parameter, this standard approach is not possible to apply. Instead, we assume that, at each stage, the only information owing the controller is that the parameter belongs to a suitable set Θ. Hence, the controller is interested in select actions directed to *minimize* the *maximum* cost generated on the corresponding set of parameters.

Specifically, in this paper we introduce a class of minimax semi-Markov control models, known as *games against nature*, to study this semi-Markov optimal control problem. The approach consists in supposing that the controller has an opponent, namely, the "nature", which, at each decision epoch, once the controller choose his/her action, the nature picks a parameter from the set Θ which might depend on the current state of the system and on the action selected. Thus, the goal of the controller is to minimize the maximum cost incurred by the nature. That is, the controller must select actions guaranteeing the best performance in the worst possible situation. Therefore, our main objective is to show the existence of minimax strategies under both, the discounted and the average criteria.

Minimax control problems have been widely studied for Markovian systems under several settings. For instance, in [1, 2, 8, 9, 14, 16] are considered particular classes of stochastic minimax control systems to study inventory and queueing models. In addition, in [28] are analyzed non-stationary minimax systems, while in [12, 21, 24] the continuous-time case is treated. A general theory for discrete-time minimax control problems is presented in [6]. On the other hand, SMCMs under discounted and average criteria typically are analyzed assuming that all components of the model are known by the controller (see, for instance, [4, 5, 10, 11, 17, 22] and their references). To the best of our knowledge, minimax semi-Markov control systems under our context have not been studied.

The paper is organized as follows. In Sect. 2 we present the semi-Markov control problem, whereas in Sect. 3 we introduce the minimax semi-Markov control model we will be dealing with, and the main assumptions. Next, in Sect. 4 we describe the minimax criteria, and Sects. 5 and 6 contain specific assumptions, preliminary results, and the main results of the discounted and average minimax criteria, respectively. Finally, in Sect. 7, we present an example of a class of semi-Markov inventory systems to illustrate our results.

2 The Semi-Markov Control Problem

Notation Given a Borel space $\mathbb{X}$ (that is, a Borel subset of a complete and separable metric space) its Borel σ-algebra is denoted by $\mathcal{B}(\mathbb{X})$, and "measurable", for either sets or functions, means "Borel measurable". Let $\mathbb{X}$ and $\mathbb{Y}$ be Borel spaces. Then a stochastic kernel $Q(dx \mid y)$ on $\mathbb{X}$ given $\mathbb{Y}$ is a function such that $Q(\cdot \mid y)$ is a probability measure on $\mathbb{X}$ for each fixed $y \in \mathbb{Y}$, and $Q(B \mid \cdot)$ is a measurable function on $\mathbb{Y}$ for each fixed $B \in \mathcal{B}(\mathbb{X})$. We denote by $\mathbb{N}$ (respectively $\mathbb{N}_0$) the set of positive (resp. nonnegative) integers; $\mathbb{R}$ (respectively $\mathbb{R}_+$) denotes the set of real (resp. nonnegative real) numbers.

Semi-Markov Control Model We consider the *semi-Markov control model*

$$SM = (\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, Q, F, D, d) \tag{1}$$

satisfying the following conditions. The *state space* $\mathbb{X}$ and the *action space* $\mathbb{A}$ are both Borel spaces. To each $x \in \mathbb{X}$, we associate a nonempty measurable subset $A(x)$ of $\mathbb{A}$, whose elements are the *admissible controls* (*or actions*) for the controller when the system is in state x. The set

$$\mathbb{K}_A = \{(x, a) : x \in \mathbb{X}, a \in A(x)\} \tag{2}$$

of *admissible state-action pairs* is assumed to be a Borel subset of $\mathbb{X} \times \mathbb{A}$. The *transition law* $Q(\cdot \mid \cdot)$ is a stochastic kernel on $\mathbb{X}$ given $\mathbb{K}_A$, and $F(\cdot \mid x, a, \theta)$ is the *distribution function of the holding time at state* $x \in \mathbb{X}$ when the control $a \in A(x)$ is chosen, which depends on an unknown and possibly non-observable parameter θ belonging to a set $\Theta(x, a)$ specified below. Finally, the *cost functions* D and d are possibly unbounded and measurable real-valued functions on $\mathbb{K}_A$.

The model (1) has the following interpretation. At time of the nth decision epoch T_n, the system is in the state $x_n = x$ and the controller chooses a control $a_n = a \in A(x)$. Then the system remains in the state x during a nonnegative random time δ_{n+1} with distribution F, and the following happens: (1) an immediate cost $D(x, a)$ is incurred; (2) the system jumps to a new state $x_{n+1} = y$ according to a transition law $Q(\cdot \mid x, a)$; and (3) a cost rate $d(x, a)$ is imposed until the transition occurs. Once the transition to state y occurs, the process is repeated.

The decision epochs are $T_n := T_{n-1} + \delta_n$ for $n \in \mathbb{N}$, and $T_0 = 0$, and the random variables $\delta_{n+1} = T_{n+1} - T_n$ are called the sojourn or holding time at state x_n.

According to the definition of the control model SM and its interpretation, the parameter may change from stage to stage. Moreover, the one-stage cost, defined in terms of the cost functions D and d, is a function $c(x, a, \theta)$, $x \in \mathbb{X}, a \in A(x), \theta \in \Theta(x, a)$. So, the optimal control problem for the controller is to find a control strategy directed to minimize his/her cost flow $\{c(x_n, a_n, \theta_n)\}_{n \in \mathbb{N}_0}$, $x_n \in \mathbb{X}$, $a_n \in A(x_n)$, $\theta_n \in \Theta(x_n, a_n)$, over an infinite horizon using either a discounted or the average cost criterion. However, since in each decision epoch the only information about the parameter θ_n is that it belongs to the set $\Theta(x_n, a_n)$, the standard procedures to solve the semi-Markov optimal control problem are not accessible to the controller.

3 Minimax Semi-Markov Control Model

In order to propose a reasonable solution, we model the semi-Markov control problem described previously as a minimax system. In this case, we assume that the controller has an opponent who selects the parameter θ in each decision epoch, which in turns determines the holding time distribution F. To fix ideas, we consider a minimax control model of the form

$$MSM = (\mathbb{X}, \mathbb{A}, \Theta, \mathbb{K}_A, \mathbb{K}, Q, F, D, d), \tag{3}$$

where $\mathbb{X}, \mathbb{A}, Q, F, D, d$, and $\mathbb{K}_A$ are as (1) and (2), and Θ is the opponent action space which is assumed to be a Borel space. The set $\mathbb{K} \in \mathcal{B}(\mathbb{X} \times \mathbb{A} \times \Theta)$ is the constraint set for the opponent. Hence, for each $(x, a) \in \mathbb{K}_A$, we suppose that

$$\Theta(x, a) := \{\theta \in \Theta : (x, a, \theta) \in \mathbb{K}\}$$

is a nonempty measurable subset of Θ representing the set of admissible actions for the opponent when the state is $x \in \mathbb{X}$ and the controller selects the action $a \in A(x)$. According to this setting, $F(\cdot|x, a, \theta)$ is the distribution function of the holding time at state $x \in \mathbb{X}$ when the control $a \in A(x)$ is chosen and the opponent selects $\theta \in \Theta(x, a)$. Furthermore, the one-stage cost c is a possible unbounded and measurable real-valued function on $\mathbb{K}$, which is defined in terms of the cost functions D and d according to the optimality criterion studied (see (5) and (12) below).

Therefore, the semi-Markov control problem can be seen as a game against the nature defined by the minimax control model (3). Indeed, once the controller selects the action $a_n = a$ when the system is in the state $x_n = x$, the opponent, the "nature", picks a parameter $\theta_n = \theta \in \Theta(x, a)$ and the system remains in the state x during the random time δ_{n+1} with distribution $F(\cdot|x, a, \theta)$. Next, the system evolves according to the model (1). Now, since the one-stage cost $c(x, a, \theta)$ depends on the parameter selected at each stage, the goal of the controller is to minimize the maximum cost incurred by the nature.

Strategies The actions or controls applied by the controller as well as his/her opponent at the decision epochs are selected according to rules known as control strategies defined as follows.

Let $\mathbb{H}_0 := \mathbb{X}$, $\mathbb{H}'_0 := \mathbb{K}_A$, and for $n \in \mathbb{N}$ let $\mathbb{H}_n := \mathbb{K}^n \times \mathbb{X}$ and $\mathbb{H}'_n := \mathbb{K}^n \times \mathbb{K}_A$. Typical elements of $\mathbb{H}_n$ and $\mathbb{H}'_n$ are written as $h_n = (x_0, a_0, \theta_0, \ldots, x_{n-1}, a_{n-1}, \theta_{n-1}, x_n)$ and $h'_n = (h_n, a_n)$, respectively. A strategy for the controller is a sequence $\pi = \{\pi_n\}$ of stochastic kernels on $\mathbb{A}$ given $\mathbb{H}_n$ such that $\pi_n(A(x_n)|h_n) = 1$ for all $h_n \in \mathbb{H}_n$ and $n \in \mathbb{N}_0$. We denote by $\Pi_{\mathbb{A}}$ the set of all strategies for the controller, and by $\mathbb{F}_{\mathbb{A}}$ the subset of all stationary strategies. As usual, every stationary strategy $\pi \in \mathbb{F}_{\mathbb{A}}$ is identified with some measurable function $f : \mathbb{X} \to \mathbb{A}$ such that $\pi_n(\cdot|h_n)$ is concentrated in $f(x_n) \in A(x_n)$ for all $h_n \in \mathbb{H}_n$ and $n \in \mathbb{N}_0$, taking the form $\pi = \{f, f, \ldots\}$. A strategy for the opponent is a sequence $\pi' = \{\pi'_n\}$ of stochastic kernels on Θ given $\mathbb{H}'_n$ such that $\pi'_n(\Theta(x_n, a_n)|h'_n) = 1$ for all $h'_n \in \mathbb{H}'_n$ and $n \in \mathbb{N}_0$. We denote by Π_Θ the set of all strategies for the opponent, and by $\mathbb{F}_\Theta \subset \Pi_\Theta$ the set of all stationary strategies. Similarly, we identify a stationary strategy $\pi' \in \mathbb{F}_\Theta$ with some measurable function $g : \mathbb{X} \times \mathbb{A} \to \Theta$ such that $\pi'_n(\cdot|h'_n)$ is concentrated in $g(x_n, a_n) \in \Theta(x_n, a_n)$ for all $h'_n \in \mathbb{H}'_n$ and $n \in \mathbb{N}_0$.

To conclude this section, we introduce the following two sets of conditions on the minimax semi-Markov model (3). The first one, Assumption 1, is a regularity condition ensuring that in a bounded time interval there are at most a finite number of transitions of the process, while in Assumption 2 we impose continuity and compactness conditions to guarantee the existence of minimax selectors.

Assumption 1 There exist $\eta > 0$ and $\varepsilon > 0$ such that

$$F(\eta \mid x, a, \theta) \leq 1 - \varepsilon \quad \forall (x, a, \theta) \in \mathbb{K}.$$

Assumption 2 (a) The cost functions $D(x, a)$ and $d(x, a)$ are lower semicontinuous (l.s.c.) on $\mathbb{K}_A$. Moreover, there exist a continuous function $W : \mathbb{X} \to [1, \infty)$ and a positive constant M_0 such that

$$\max\{D(x, a), d(x, a)\} \leq M_0 W(x) \quad \forall (x, a) \in \mathbb{K}_A.$$

(b) The transition law Q is weakly continuous, that is, for each continuous and bounded function $u : \mathbb{X} \to \mathbb{R}$, the function

$$(x, a) \longmapsto \int_{\mathbb{X}} u(y)\, Q(dy \mid x, a)$$

is continuous on $\mathbb{K}_A$.

(c) The function

$$(x, a) \longmapsto \int_{\mathbb{X}} W(y)\, Q(dy \mid x, a)$$

is continuous on $\mathbb{K}_A$.

(d) The multifunction $x \to A(x)$ is upper semi-continuous (u.s.c.), and the set $A(x)$ is compact for each $x \in \mathbb{X}$.

(e) The multifunction $(x, a) \to \Theta(x, a)$ is l.s.c., and for each $(x, a) \in \mathbb{K}_A$, $\Theta(x, a)$ is a σ-compact set.

We denote by $\mathbb{B}_W$ the normed linear space of all measurable functions $u : \mathbb{X} \to \mathbb{R}$ with norm

$$\|u\|_W := \sup_{x\in\mathbb{X}} \frac{|u(x)|}{W(x)} < \infty, \tag{4}$$

and by $\mathbb{L}_W$ the subspace of l.s.c. functions in $\mathbb{B}_W$.

Remark 1 (a) As is well-known, the Assumption 2(b) can be substituted by the following equivalent condition: For each l.s.c. and bounded below function $u : \mathbb{X} \to \mathbb{R}$, the function

$$(x, a) \longmapsto \int_{\mathbb{X}} u(y)\, Q(dy \mid x, a)$$

is l.s.c. on $\mathbb{K}_A$.

(b) It is easy to prove that $\mathbb{L}_W$ is a closed subset of $\mathbb{B}_W$. Hence, due to $\mathbb{B}_W$ is a Banach space, we have that $\mathbb{L}_W$ is a complete subspace of $\mathbb{B}_W$.

4 Minimax Criteria

To study the discounted criterion, we assume that the costs are continuously discounted. That is, for a discount factor $\alpha > 0$, a cost C incurred at time t is equivalent to a cost $C \exp(-\alpha t)$ at time 0. Then we define the one-stage cost for the discounted criterion as

$$c(x, a, \theta) := D(x, a) + d(x, a) \int_0^\infty \int_0^t \exp(-\alpha s) ds\, F(dt \mid x, a, \theta), \quad (x, a, \theta) \in \mathbb{K}. \tag{5}$$

Furthermore, for $(x, a, \theta) \in \mathbb{K}$, let

$$\beta_\alpha(x, a, \theta) := \int_0^\infty \exp(-\alpha s) F(ds \mid x, a, \theta) \tag{6}$$

and

$$\tau_\alpha(x, a, \theta) := \frac{1 - \beta_\alpha(x, a, \theta)}{\alpha}.$$

Hence, it is easy to see that the one-stage cost (5) takes the form

$$c(x, a, \theta) = D(x, a) + \tau_\alpha(x, a, \theta) d(x, a), \quad (x, a, \theta) \in \mathbb{K}. \tag{7}$$

For a fixed $\alpha > 0$, and for each pair of strategies $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_{\Theta}$ and initial state $x \in \mathbb{X}$, we define the total expected discounted cost as

$$V(x, \pi, \pi') := E_x^{\pi\pi'} \left[\sum_{n=0}^{\infty} \exp(-\alpha T_n) c(x_n, a_n, \theta_n) \right], \tag{8}$$

where $E_x^{\pi\pi'}$ denotes the expectation operator with respect to the probability measure $P_x^{\pi\pi'}$ induced by $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_{\Theta}$, given $x_0 = x$ (for the construction of $P_x^{\pi\pi'}$ see, for instance, [3]).

Let

$$V'(x, \pi) := \sup_{\pi' \in \Pi_{\Theta}} V(x, \pi, \pi'), \quad x \in \mathbb{X}, \ \pi \in \Pi_{\mathbb{A}}.$$

Then, the discounted minimax semi-Markov control problem to the controller is to find a strategy $\pi^* \in \Pi_{\mathbb{A}}$ such that

$$V'(x, \pi^*) = \inf_{\pi \in \Pi_{\mathbb{A}}} V'(x, \pi) = \inf_{\pi \in \Pi_{\mathbb{A}}} \sup_{\pi' \in \Pi_{\Theta}} V(x, \pi, \pi') =: V^*(x), \quad x \in \mathbb{X}. \tag{9}$$

In this case, the strategy π^* is said to be discounted minimax, whereas V^* is the discounted optimal value function.

On the other hand, to define the average cost criterion, we first define the *mean holding time* in state $x \in \mathbb{X}$, when the controller and the opponent select $a \in A(x)$ and $\theta \in \Theta(x, a)$, respectively, as

$$\tau(x, a, \theta) := \int_0^{\infty} t F(dt \mid x, a, \theta), \quad (x, a, \theta) \in \mathbb{K}. \tag{10}$$

Then, for each $x \in \mathbb{X}$ and $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_{\Theta}$, we define the *long-run expected average cost* by

$$J(x, \pi, \pi') := \limsup_{n \to \infty} \frac{E_x^{\pi\pi'}[\sum_{k=0}^{n-1} c(x_k, a_k, \theta_k)]}{E_x^{\pi\pi'}[\sum_{k=0}^{n-1} \tau(x_k, a_k, \theta_k)]}, \tag{11}$$

where

$$c(x, a, \theta) := D(x, a) + \tau(x, a, \theta) d(x, a), \quad (x, a, \theta) \in \mathbb{K}, \tag{12}$$

is the one-stage cost for the average criterion.

Defining

$$J'(x, \pi) := \sup_{\pi' \in \Pi_{\Theta}} J(x, \pi, \pi'), \quad x \in \mathbb{X}, \ \pi \in \Pi_{\mathbb{A}}, \tag{13}$$

the average minimax semi-Markov control problem to the controller is to find a strategy $\pi^* \in \Pi_{\mathbb{A}}$ such that

$$J'(x, \pi^*) = \inf_{\pi \in \Pi_{\mathbb{A}}} J'(x, \pi) = \inf_{\pi \in \Pi_{\mathbb{A}}} \sup_{\pi' \in \Pi_{\Theta}} J(x, \pi, \pi') := J^*(x), \quad x \in \mathbb{X}. \tag{14}$$

The strategy π^* is said to be average minimax and J^* is the average optimal value function.

Our objective is to show the existence of discounted and average minimax strategies for the semi-Markov model (3).

5 Discounted Cost Criterion

In this section we analyze the minimax discounted cost criterion defined by (5)–(9).

From (6), using properties of the conditional expectation, we can write the performance index (8) as

$$V(x,\pi,\pi') = E_x^{\pi\pi'}\left[c(x_0,a_0,\theta_0) + \sum_{n=1}^{\infty}\prod_{k=0}^{n-1}\beta_\alpha(x_k,a_k,\theta_k)c(x_n,a_n,\theta_n)\right],$$
$$x\in\mathbb{X},\ (\pi,\pi')\in\Pi_{\mathbb{A}}\times\Pi_\Theta. \tag{15}$$

A first consequence of Assumptions 1 and 2 is the following.

Lemma 1 *Under Assumptions* 1 *and* 2(a) *we have*:

(a) $\rho_\alpha := \sup_{(x,a,\theta)\in\mathbb{K}}\beta_\alpha(x,a,\theta) < 1$;
(b) *For all* $(x,a,\theta)\in\mathbb{K}$, *and some constant* $M_1>0$,

$$|c(x,a,\theta)|\le M_1W(x).$$

Proof (a) Integrating by parts in (6), from Assumption 1, for each $(x,a,\theta)\in\mathbb{K}$,

$$\begin{aligned}\beta_\alpha(x,a,\theta) &:= \int_0^\infty \exp(-\alpha t)F(dt\mid x,a,\theta) = -\int_0^\infty F(t\mid x,a,\theta)d\left(\exp(-\alpha t)\right)\\ &= \alpha\left\{\int_0^\eta \exp(-\alpha t)F(t\mid x,a,\theta)dt\right.\\ &\quad\left.+\int_\eta^\infty \exp(-\alpha t)F(t\mid x,a,\theta)dt\right\}\\ &\le (1-\varepsilon)\left(1-e^{-\alpha\eta}\right)+e^{-\alpha\eta}<1,\quad 0<\eta<\infty.\end{aligned}$$

Hence, $\rho_\alpha := \sup_{(x,a,\theta)\in\mathbb{K}}\beta_\alpha(x,a,\theta) < 1$.

(b) Observe that from part (a), $\tau_\alpha(x,a,\theta)\le 1/\alpha$ for all $(x,a,\theta)\in\mathbb{K}$. Then, from Assumption 2(a), part (b) follows easily from (7) with $M_1 := (1+1/\alpha)M_0$. □

Note that Assumption 2 (see Lemma 1(b)) allows an unbounded one-stage cost function $c(x,a,\theta)$ provided that it is majorized by the "bounding" function W. However, to state our main result, we need to impose a condition to limit its growth. Moreover, we need a continuity condition on the holding time distribution.

Assumption 3 (a) There exists a positive constant b such that

$$1\le b<\rho_\alpha^{-1},$$

and for all $(x,a)\in\mathbb{K}_A$

$$\int_{\mathbb{X}} W(y)Q(dy\mid x,a)\le bW(x), \tag{16}$$

where W is the function in Assumption 2.

(b) For each $t\ge 0$, $F(t\mid x,a,\theta)$ is continuous in $\mathbb{K}$.

Remark 2 (a) Observe that for all $x \in \mathbb{X}$, $n \in \mathbb{N}_0$, and $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_{\Theta}$, (16) yields

$$E_x^{\pi\pi'}\left[W(x_{n+1})\right] \leq b E_x^{\pi\pi'}\left[W(x_n)\right],$$

which in turns implies

$$E_x^{\pi\pi'}\left[W(x_{n+1})\right] \leq b^n W(x). \tag{17}$$

Therefore, from (15) and Lemma 1, for each $x \in \mathbb{X}$, and $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_{\Theta}$,

$$\begin{aligned} \left|V(x, \pi, \pi')\right| &\leq E_x^{\pi\pi'}\left|c(x_0, a_0, \theta_0) + \sum_{n=1}^{\infty} \rho_\alpha^n c(x_n, a_n, \theta_n)\right| \\ &\leq M_1 W(x) \sum_{n=0}^{\infty} (b\rho_\alpha)^n = \frac{M_1 W(x)}{1 - b\rho_\alpha}. \end{aligned}$$

Then,

$$\|V^*\|_W \leq \frac{M_1}{1 - b\rho_\alpha}. \tag{18}$$

(b) From Assumption 3(b), the function $\beta_\alpha(x, a, \theta)$ in (6) is continuous on $\mathbb{K}$, and in addition, from Assumption 2(a) the cost function $c(x, a, \theta)$ at (7) is l.s.c. on $\mathbb{K}_A$ and continuous in θ.

For $u \in \mathbb{B}_W$ and $(x, a, \theta) \in \mathbb{K}$, we define

$$H_\alpha(u, x, a, \theta) := c(x, a, \theta) + \beta_\alpha(x, a, \theta) \int_{\mathbb{X}} u(y) Q(dy \mid x, a)$$

and

$$T_\alpha u(x) := \inf_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} H_\alpha(u, x, a, \theta). \tag{19}$$

Lemma 2 *If Assumptions 1, 2, and 3 hold, then:*

(a) *The operator T_α is a contraction with modulus $b\rho_\alpha < 1$.*
(b) *T_α maps $\mathbb{L}_W$ into itself.*
(c) *For each $u \in \mathbb{L}_W$, there exists $f^* \in \mathbb{F}_{\mathbb{A}}$ such that*

$$T_\alpha u(x) = \sup_{\theta \in \Theta(x, f^*)} H_\alpha(u, x, f^*, \theta), \quad x \in \mathbb{X}.$$

Proof (a) Let $u, u' \in \mathbb{B}_W$. Then, from definition of the norm $\|\cdot\|_W$ in (4), Lemma 1(a) and Assumption 3(a), we have

$$\begin{aligned} H_\alpha(u, x, a, \theta) &\leq H_\alpha(u', x, a, \theta) + \beta_\alpha(x, a, \theta) \int_{\mathbb{X}} \left|u(y) - u'(y)\right| Q(dy \mid x, a) \\ &\leq H_\alpha(u', x, a, \theta) + b\rho_\alpha \left\|u - u'\right\|_W W(x) \quad \forall (x, a, \theta) \in \mathbb{K}, \end{aligned}$$

which implies (see (19))

$$T_\alpha u(x) \leq T_\alpha u'(x) + b\rho_\alpha \|u - u'\|_W W(x) \quad \forall x \in \mathbb{X}.$$

Thus,

$$T_\alpha u(x) - T_\alpha u'(x) \leq b\rho_\alpha \left\| u - u' \right\|_W W(x) \quad \forall x \in \mathbb{X}. \tag{20}$$

A similar argument shows that

$$T_\alpha u'(x) - T_\alpha u(x) \leq b\rho_\alpha \left\| u - u' \right\|_W W(x) \quad \forall x \in \mathbb{X}. \tag{21}$$

Hence, combining (20) and (21) we obtain

$$\left\| T_\alpha u - T_\alpha u' \right\|_W \leq b\rho_\alpha \left\| u - u' \right\|_W,$$

which proves the part (a).

(b) Let $u \in \mathbb{L}_W$, and we define

$$\hat{H}(u, x, a) := \sup_{\theta \in \Theta(x,a)} H_\alpha(u, x, a, \theta), \quad (x, a) \in \mathbb{K}_A.$$

To prove the part (b) it is sufficient to show:

(i) $|\hat{H}(u, x, a)| \leq \hat{M} W(x)$, for all $(x, a) \in \mathbb{K}_A$ and some constant $\hat{M} < \infty$; and
(ii) $\hat{H}$ is l.s.c. on $\mathbb{K}_A$.

Indeed, from (19), the relation (i) yields

$$|T_\alpha u(x)| \leq \hat{M} W(x) \quad \forall x \in \mathbb{X},$$

which implies

$$\|T_\alpha u\|_W < \infty. \tag{22}$$

Now, if (ii) holds, from (i), we have that the function

$$(x, a) \mapsto \hat{H}(u, x, a) + \hat{M} W(x) \tag{23}$$

is nonnegative and l.s.c. on $\mathbb{K}_A$. Hence, from Assumption 2(d) and due to a well-known result is [25], (see also [7], Proposition D.5), we deduce that

$$\begin{aligned} H(u, x) &:= \inf_{a \in A(x)} \left\{ \hat{H}(u, x, a) + \hat{M} W(x) \right\} \\ &= T_\alpha u(x) + \hat{M} W(x) \end{aligned} \tag{24}$$

is a l.s.c. function on $\mathbb{X}$. Then, by the continuity of $W(\cdot)$, we obtain that

$$T_\alpha u(x) = H(u, x) - \hat{M} W(x)$$

is a l.s.c. function on $\mathbb{X}$. This fact together (22) implies that T_α maps $\mathbb{L}_W$ into itself.

To conclude the proof of part (b), we prove the points (i) and (ii).

Let $u \in \mathbb{L}_W$. From (4), Lemma 1(b) and Assumption 3(a), the point (i) easily follows with $\hat{M} := M_1 + b\rho_\alpha \|u\|_W$.

On the other hand, observe that

$$|u(x)| \leq \|u\|_W W(x), \quad x \in \mathbb{X}.$$

Thus, by the continuity of $W(\cdot)$, the function

$$v\,(x) := u\,(x) + \|u\|_W \, W\,(x)\,,$$

is nonnegative and l.s.c., which implies that (see Remark 1(a)) $\int_{\mathbb{X}} v(y)Q(dy \mid x, a)$ is a l.s.c. function on $\mathbb{K}_A$. Hence, from Assumption 2(c),

$$\int_{\mathbb{X}} u(y)Q(dy \mid x, a) = \int_{\mathbb{X}} v(y)Q(dy \mid x, a) - \|u\|_W \int_{\mathbb{X}} W(y)Q(dy \mid x, a)$$

is l.s.c. on $\mathbb{K}_A$. Therefore (see Remark 2(b)), $H_\alpha(u, x, a, \theta)$ is l.s.c. on $\mathbb{K}$.

Now, let $\{(x_j, a_j)\} \subset \mathbb{K}_A$ be a sequence converging to $(x, a) \in \mathbb{K}_A$, and $\theta \in \Theta(x, a)$ be arbitrary. Then, since $\Theta(x, a)$ is σ-compact (see Assumption 2(e)), there exists $\theta_j \in \Theta(x_j, a_j)$ such that $\theta_j \to \theta$. Hence,

$$\begin{aligned}
\liminf_{j\to\infty} \hat{H}(u, x_j, a_j) &= \liminf_{j\to\infty} \sup_{\theta\in\Theta(x_j,a_j)} H_\alpha(u, x_j, a_j, \theta) \\
&\geq \liminf_{j\to\infty} H_\alpha(u, x_j, a_j, \theta_j) \\
&\geq H_\alpha(u, x, a, \theta),
\end{aligned}$$

where the last inequality is from the lower semi-continuity of H_α. Since θ was arbitrarily chosen, we have

$$\liminf_{j\to\infty} \hat{H}(u, x_j, a_j) \geq \hat{H}(u, x, a),$$

which implies that $\hat{H}$ is l.s.c. on $\mathbb{K}_A$. This completes the proof of part (b).

(c) From the nonnegativity and lower semi-continuity of the function defined in (23), and Assumption 2(d), standard arguments on the existence of minimizers (see, for instance, [23, 25]) imply that there exists $f^* \in \mathbb{F}_A$ such that

$$\inf_{a\in A(x)} \left\{ \hat{H}(u, x, a) + \hat{M}W\,(x) \right\} = \hat{H}(u, x, f^*) + \hat{M}W\,(x)\,.$$

Thus, combining this relation with (24) we obtain the part (c). □

Remark 3 Since T_α is a contraction operator which maps $\mathbb{L}_W$ into itself (see Lemma 2) and $\mathbb{L}_W \subset \mathbb{B}_W$ is complete (see Remark 1(b)), from Banach's Fixed Point Theorem, there exists a unique function $\tilde{u} \in \mathbb{L}_W$ such that for all $x \in \mathbb{X}$,

$$\tilde{u}\,(x) = T_\alpha \tilde{u}\,(x)\,, \tag{25}$$

and

$$\left\| T_\alpha^n u - \tilde{u} \right\|_W \leq (b\rho_\alpha)^n \, \|u - \tilde{u}\|_W \quad \forall u \in \mathbb{L}_W,\; n \in \mathbb{N}_0. \tag{26}$$

For each $n \in \mathbb{N}$, $x \in \mathbb{X}$, and $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_\Theta$, we define the n-stage expected discounted cost as (see (15))

$$V^n(x, \pi, \pi') := \begin{cases} E_x^{\pi\pi'}[c(x_0, a_0, \theta_0)], & n = 1 \\ E_x^{\pi\pi'}[c(x_0, a_0, \theta_0) + \sum_{j=1}^{n-1} \prod_{k=0}^{j-1} \beta_\alpha(x_k, a_k, \theta_k) c(x_j, a_j, \theta_j)], & n \geq 2. \end{cases}$$

In addition, we define the sequence $\{v_n\}$ in $\mathbb{L}_W$ as $v_0 = 0$, and for $n \in \mathbb{N}$,

$$v_n(x) = T_\alpha v_{n-1}(x), \quad x \in \mathbb{X}.$$

Now, we state our first main result as follows

Theorem 1 *If Assumptions 1, 2 and 3 hold, then*:

(a) *The optimal value function* (9) *is the unique solution in* $\mathbb{L}_W$ *satisfying*

$$V^*(x) = T_\alpha V^*(x), \quad x \in \mathbb{X}.$$

(b) *For each* $n \in \mathbb{N}$,

$$\left\| v_n - V^* \right\|_W \leq M_1 \frac{(b\rho_\alpha)^n}{1 - b\rho_\alpha}.$$

(c) *There exists* $f^* \in \mathbb{F}_\mathbb{A}$ *such that*

$$V^*(x) = \sup_{\theta \in \Theta(x, f^*(x))} \left\{ c(x, f^*(x), \theta) + \beta_\alpha(x, f^*(x), \theta) \int_\mathbb{X} V^*(y) Q(dy \mid x, f^*(x)) \right\},$$

and moreover f^* *is a discounted cost minimax strategy for the controller, that is,*

$$V^*(x) = \sup_{\pi' \in \Pi_\Theta} V(x, f^*, \pi').$$

Proof (a)–(b) First observe that $v_n = T_\alpha^n v_0$ for all $n \in \mathbb{N}_0$. Then, taking $u = v_0$ in (26) we have

$$\|v_n - \tilde{u}\|_W \leq (b\rho_\alpha)^n \|\tilde{u}\|_W \quad \forall n \in \mathbb{N}_0.$$

Therefore, from (18) and (25), the parts (a) and (b) of the theorem will be proved if we show that $\tilde{u} = V^*$.

Let $f \in \mathbb{F}_\mathbb{A}$ be a selector such that (see Lemma 2(c))

$$\tilde{u}(x) = \sup_{\theta \in \Theta(x, f(x))} H_\alpha(\tilde{u}, x, f(x), \theta), \quad x \in \mathbb{X}.$$

Then,

$$\tilde{u}(x) \geq c(x, f, \theta) + \beta_\alpha(x, f, \theta) \int_\mathbb{X} \tilde{u}(y) Q(dy \mid x, f) \quad \forall x \in \mathbb{X}, \theta \in \Theta(x, f(x)). \tag{27}$$

Now, let $\pi' \in \Pi_\Theta$ be an arbitrary strategy for the opponent and $\{(x_n, f(x_n), \theta_n)\}$ be a sequence of state-actions corresponding to application of the strategies f and π'. Then, iterating inequality (27), a straight forward calculation yields,

$$\tilde{u}(x) \geq E_x^{f\pi'} \left[c(x_0, f, \theta_0) + \sum_{j=1}^{n-1} \prod_{k=0}^{j-1} \beta_\alpha(x_k, f, \theta_k) c(x_j, f, \theta_j) \right]$$
$$+ \prod_{k=0}^{n-1} \beta_\alpha(x_k, f, \theta_k) E_x^{f\pi'} \left[\tilde{u}(x_n) \right].$$

That is,

$$\tilde{u}(x) \geq V^n\left(x, f, \pi'\right) + \prod_{k=0}^{n-1} \beta_\alpha(x_k, f, \theta_k) E_x^{f\pi'}\left[\tilde{u}(x_n)\right] \quad \forall x \in \mathbb{X}. \tag{28}$$

Observe that from (17), we have that

$$\prod_{k=0}^{n-1} \beta_\alpha(x_k, f, \theta_k) E_x^{f\pi'}\left[\left|\tilde{u}(x_n)\right|\right] \leq (b\rho_\alpha)^n \|\tilde{u}\|_W W(x), \quad x \in \mathbb{X}.$$

Hence, letting $n \to \infty$ in (28) we obtain

$$\tilde{u}(x) \geq V\left(x, f, \pi'\right) \quad \forall x \in \mathbb{X}, \pi' \in \Pi_\Theta, \tag{29}$$

which implies

$$\tilde{u}(x) \geq \sup_{\pi' \in \Pi_\Theta} V\left(x, f, \pi'\right) \quad \forall x \in \mathbb{X}. \tag{30}$$

Therefore,

$$\tilde{u}(x) \geq V^*(x) \quad \forall x \in \mathbb{X}. \tag{31}$$

On the other hand, since the functions c and β_α are continuous in θ (see Remark 2(b)), from (25), Assumption 2(e), and by applying a suitable Measurable Selection Theorem (see [23]), for every $\varepsilon > 0$, there exists $g : \mathbb{K}_\mathbb{A} \to \Theta$, with $g(x,a) \in \Theta(x,a)$, such that

$$\begin{aligned} \tilde{u}(x) &= \inf_{a \in A(x)} \left\{ c(x,a,g) + \beta_\alpha(x,a,g) \int_\mathbb{X} \tilde{u}(y) Q(dy \mid x,a) \right\} + \varepsilon \\ &\leq c(x,a,g) + \beta_\alpha(x,a,g) \int_\mathbb{X} \tilde{u}(y) Q(dy \mid x,a) + \varepsilon \quad \forall x \in \mathbb{X}, a \in A(x). \end{aligned} \tag{32}$$

Let $\pi \in \Pi_\mathbb{A}$ be an arbitrary strategy. Then, similarly as in (28), iterating (32) we get

$$\tilde{u}(x) \leq V^n(x, \pi, g) + \prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, g) E_x^{\pi g}\left[\tilde{u}(x_n)\right] + \frac{\varepsilon}{1-\rho_\alpha} \quad \forall x \in \mathbb{X}.$$

Thus, (see (29)), letting $n \to \infty$ we have

$$\tilde{u}(x) \leq V(x, \pi, g). \tag{33}$$

Then, since

$$V(x, \pi, g) \leq \sup_{g \in \mathbb{F}_\Theta} V(x, \pi, g) \leq \sup_{\pi' \in \Pi_\Theta} V\left(x, \pi, \pi'\right) \quad \forall x \in \mathbb{X},$$

and $\pi \in \Pi_\mathbb{A}$ is arbitrary, (33) implies

$$\tilde{u}(x) \leq \inf_{\pi \in \Pi_\mathbb{A}} \sup_{\pi' \in \Pi_\Theta} V(x, \pi, \pi') = V^*(x) \quad \forall x \in \mathbb{X},$$

which combined with (31) yield $\tilde{u} = V^*$.

(c) The existence of $f^* \in \mathbb{F}_{\mathbb{A}}$ follows from the part (a) and Lemma 2(c). In addition, applying similar arguments as the proof of part (a) (see (30)) we have

$$V^*(x) := \inf_{\pi\in\Pi_{\mathbb{A}}} \sup_{\pi'\in\Pi_\Theta} V\left(x,\pi,\pi'\right) \geq \sup_{\pi'\in\Pi_\Theta} V\left(x,f^*,\pi'\right) \quad \forall x\in\mathbb{X}.$$

Therefore

$$V^*(x) = \sup_{\pi'\in\Pi_\Theta} V(x,f^*,\pi') \quad \forall x\in\mathbb{X},$$

that is, f^* is a minimax strategy. □

6 Average Cost Criterion

To analyze the average minimax semi-Markov control problem defined by (10)–(14), we need to impose continuity and boundedness conditions on the mean holding time. That is, it is necessary the existence of positive constant m and M satisfying

$$m \leq \tau(x,a,\theta) \leq M \quad \forall (x,a,\theta)\in\mathbb{K}. \tag{34}$$

The first inequality in (34) is a consequence of Assumption 1 because for all $(x,a,\theta)\in\mathbb{K}$,

$$\tau(x,a,\theta) \geq \int_\eta^\infty tF(dt \mid x,a,\theta) \geq \eta\varepsilon =: m.$$

Hence,

$$\inf_{(x,a,\theta)\in\mathbb{K}} \tau(x,a,\theta) \geq m. \tag{35}$$

The second inequality together the continuity condition is given in the next assumption.

Assumption 4 (a) There exists a positive constant M such that for all $(x,a,\theta)\in\mathbb{K}$,

$$\tau(x,a,\theta) \leq M.$$

(b) The function $\tau(x,a,\theta)$ is continuous on $\mathbb{K}$.

Observe that from Assumptions 2(a) and 4(a), the one-stage cost function (12) satisfies, for some constant $M_2 > 0$,

$$|c(x,a,\theta)| \leq M_2 W(x) \quad \forall (x,a,\theta)\in\mathbb{K}. \tag{36}$$

Furthermore, because the asymptotic analysis for the average criterion, we need to impose the following ergodicity condition.

Assumption 5 There exist a probability measure μ on $\mathbb{X}$, an u.s.c. function $\phi : \mathbb{K}_A \to [0,1]$, and a constant $\gamma\in(0,1)$, such that

(a) $Q(D \mid x,a) \geq \phi(x,a)\mu(D)$, for all $(x,a)\in\mathbb{K}_A$ and $D\in\mathcal{B}(\mathbb{X})$.

(b) $\int_{\mathbb{X}} W(y)Q(dy \mid x, a) \leq \gamma W(x) + \phi\,(x, a) \int_{\mathbb{X}} W(y)\mu(dy)$, where

$$\mu(W) := \int_{\mathbb{X}} W\,(y)\,\mu\,(dy) < \infty.$$

(c) $\int_{\mathbb{X}} \phi(x, f(x))\mu(dx) > 0$ for each $f \in \mathbb{F}_{\mathbb{A}}$.

An important consequence of Assumption 5 is that for each $f \in \mathbb{F}_A$, the Markov chain defined by $Q(\cdot \mid \cdot, f)$ is μ-irreducible and *positive Harris recurrent*, which is proved in [27].

Let $f \in \mathbb{F}_A$, and $Q^n(\cdot \mid \cdot, f)$ the n-step transition kernel associated to f. Then, by Assumption 5(b),

$$\int_{\mathbb{X}} W\,(y)\,Q^n\,(dy \mid x, f) \leq \gamma^n W(x) + (1 + \gamma + \cdots + \gamma^{n-1})\mu(W) \leq \left(1 + \frac{\mu(W)}{1-\gamma}\right) W(x),$$

and this implies that for all function $v \in \mathbb{B}_W$,

$$\lim_{n\to\infty} \frac{1}{n} \int_{\mathbb{X}} v\,(y)\,Q^n\,(dy \mid x, f) = 0. \tag{37}$$

The main result for the average criterion is stated as follows.

Theorem 2 *Suppose that Assumptions 1, 2, 4 and 5 hold. Then, there exist a constant j^*, a function $h^* \in \mathbb{L}_W$, and a stationary strategy $f^* \in \mathbb{F}_{\mathbb{A}}$ such that for all $x \in \mathbb{X}$,*

$$\begin{aligned} h^*(x) &= \inf_{a\in A(x)} \sup_{\theta\in\Theta(x,a)} \left\{ c(x, a, \theta) - j^*\tau\,(x, a, \theta) + \int_{\mathbb{X}} h^*(y)Q(dy \mid x, a) \right\} \\ &= \sup_{\theta\in\Theta(x, f^*(x))} \left\{ c(x, f^*(x), \theta) - j^*\tau(x, f^*(x), \theta) + \int_{\mathbb{X}} h^*(y)Q(dy \mid x, f^*(x)) \right\}. \end{aligned}$$

Furthermore, j^* is the optimal average cost, and f^* is an average minimax strategy, that is, for all $x \in \mathbb{X}$,

$$\begin{aligned} j^* &= \sup_{\pi'\in\Pi_\Theta} J(x, f^*, \pi') \\ &= \inf_{\pi\in\Pi_{\mathbb{A}}} \sup_{\pi'\in\Pi_\Theta} J(x, \pi, \pi'). \end{aligned}$$

The proof of the Theorem 2 is based in a data transformation introduced in [26] and applied by several authors to semi-Markov control models (see, for instance, [4, 5]). We introduce this procedure in our context as follows.

Let τ be a real number such that

$$0 < \tau < m, \tag{38}$$

(see (35)). Define the function $\hat{c} : \mathbb{K} \to \mathbb{R}$ and the stochastic kernel $\hat{Q}$ on $\mathbb{X}$ given $\mathbb{K}$ by

$$\hat{c}\,(x, a, \theta) := \frac{c\,(x, a, \theta)}{\tau\,(x, a, \theta)}, \tag{39}$$

and

$$\hat{Q}(B|x,a,\theta) := \frac{\tau}{\tau\,(x,a,\theta)} Q(B|x,a) + \left(1 - \frac{\tau}{\tau\,(x,a,\theta)}\right)\delta_x(B), \tag{40}$$

where $\delta_x(\cdot)$ is the Dirac measure concentrated at x.

Then, the cost function $\hat{c}$ and the stochastic kernel $\hat{Q}$ define the minimax Markov control model

$$(\mathbb{X}, \mathbb{A}, \Theta, \mathbb{K}_A, \mathbb{K}, \hat{Q}, \hat{c}). \tag{41}$$

Hence, our approach consists in to analyze the minimax control problem corresponding to the model (41) whose solution proves the Theorem 2. To this end, we need the following results.

Lemma 3 *If Assumptions 1, 2, and 4 hold, then Assumption 2(b), (c) and relation (36) are satisfied when c, Q, and $\mathbb{K}_A$ are replaced by $\hat{c}$, $\hat{Q}$, and $\mathbb{K}$, respectively.*

Proof First we will prove that the transition law $\hat{Q}$ is weakly continuous. Let $u : \mathbb{X} \to \mathbb{R}$ be a continuous and bounded function, then from (40) we have

$$\int_{\mathbb{X}} u\,(y)\,\hat{Q}\,(dy \mid x,a,\theta) = \frac{\tau}{\tau\,(x,a,\theta)} \int_{\mathbb{X}} u\,(y)\,Q(dy|x,a) + \left(1 - \frac{\tau}{\tau\,(x,a,\theta)}\right) u(x).$$

Thus, by (35) and Assumptions 2(a) and 4(b), we conclude that the function

$$(x,a,\theta) \mapsto \int_{\mathbb{X}} u\,(y)\,\hat{Q}\,(dy \mid x,a,\theta)$$

is continuous on $\mathbb{K}$. Similarly, we prove that the function

$$(x,a,\theta) \mapsto \int_{\mathbb{X}} W\,(y)\,\hat{Q}\,(dy \mid x,a,\theta)$$

is continuous on $\mathbb{K}$.

Finally, from (35) and (36),

$$\left|\hat{c}\,(x,a,\theta)\right| = \frac{|c\,(x,a,\theta)|}{\tau\,(x,a,\theta)} \le M_* W\,(x) \quad \forall\,(x,a,\theta) \in \mathbb{K},$$

where $M_* := \frac{M_2}{m}$. □

The model (41) satisfies Assumption 5, `mutatis mutandi`, according to the following result.

Lemma 4 *Under Assumptions 1, 2, 4, and 5, there exist a probability measure μ on $\mathbb{X}$, an u.s.c. function $\hat{\phi} : \mathbb{K} \to [0,1]$, and a constant $\hat{\gamma} \in (0,1)$, such that*

(a) $\hat{Q}(D \mid x,a,\theta) \ge \hat{\phi}\,(x,a,\theta)\mu(D)$, *for all* $(x,a,\theta) \in \mathbb{K}$ *and* $D \in \mathcal{B}(\mathbb{X})$.
(b) $\int_{\mathbb{X}} W(y)\hat{Q}(dy \mid x,a,\theta) \le \hat{\gamma} W(x) + \hat{\phi}(x,a,\theta) \int_{\mathbb{X}} W(y)\mu(dy)$, *where*

$$\int_{\mathbb{X}} W\,(y)\,\mu\,(dy) < \infty.$$

(c) $\int_{\mathbb{X}} \hat{\phi}(x, f(x), \theta)\mu(dx) > 0$ *for each* $f \in \mathbb{F}_{\mathbb{A}}$.

Proof Let

$$\hat{\phi}(x, a, \theta) := \frac{\tau}{\tau(x, a, \theta)}\phi(x, a), \quad (x, a, \theta) \in \mathbb{K},$$

$$\hat{\gamma} := 1 - \frac{\tau}{M}(1 - \gamma),$$

and we consider the probability measure μ in Assumption 5.

Observe that from (38) and Assumption 4, $\hat{\phi} : \mathbb{K} \to [0, 1]$ and $0 < \hat{\gamma} < 1$ hold true. Therefore:

(a) Assumption 5(a) yields

$$\hat{Q}(D \mid x, a, \theta) \geq \hat{\phi}(x, a, \theta)\mu(D) \quad \forall (x, a, \theta) \in \mathbb{K}, \ D \in \mathcal{B}(\mathbb{X});$$

(b) from Assumption 5(b) we obtain

$$\int_{\mathbb{X}} W(y)\hat{Q}(dy \mid x, a, \theta) \leq \hat{\gamma} W(x) + \hat{\phi}(x, a, \theta)\int_{\mathbb{X}} W(y)\mu(dy);$$

and finally,

(c) from Assumption 5(c) we conclude that

$$\int_{\mathbb{X}} \hat{\phi}(x, f(x), \theta)\mu(dx) > 0.$$

□

As a consequence of Lemma 4, we can state the following result for the minimax Markov model (41) borrowed from Theorem 5.2 in [6], (see also [13, 14, 16]).

Proposition 1 *Under Assumptions 1, 2, 4, and 5, there exist a constant* $\hat{j}$, *a function* $\hat{h} \in \mathbb{L}_W$, *and a stationary policy* $\hat{f} \in \mathbb{F}_{\mathbb{A}}$ *such that for all* $x \in \mathbb{X}$,

$$\hat{j} + \hat{h}(x) = \inf_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} \left\{ \hat{c}(x, a, \theta) + \int_{\mathbb{X}} \hat{h}(y)\hat{Q}(dy \mid x, a, \theta) \right\} \tag{42}$$

$$= \sup_{\theta \in \Theta(x,\hat{f})} \left\{ \hat{c}(x, \hat{f}, \theta) + \int_{\mathbb{X}} \hat{h}(y)\hat{Q}(dy \mid x, \hat{f}, \theta) \right\}. \tag{43}$$

6.1 Proof of Theorem 2

From (42), we have that for each $(x, a) \in \mathbb{K}_A$,

$$\hat{j} + \hat{h}(x) \leq \sup_{\theta \in \Theta(x,a)} \left\{ \hat{c}(x, a, \theta) + \int_{\mathbb{X}} \hat{h}(y)\hat{Q}(dy \mid x, a, \theta) \right\}.$$

Observe that from Assumption 3(b) and (40), the function

$$\theta \longmapsto \int_{\mathbb{X}} \hat{h}(y)\hat{Q}(dy \mid x, a, \theta)$$

is continuous for all $(x, a) \in \mathbb{K}_A$. Then, for every $\epsilon > 0$, from Assumption 2(e) and by applying a suitable Measurable Selection Theorem, there exists $g \in \mathbb{F}_\Theta$ such that

$$\hat{j} + \hat{h}(x) \leq \hat{c}(x, a, g) + \int_{\mathbb{X}} \hat{h}(y) \hat{Q}(dy \mid x, a, g) + \epsilon/M \quad \forall (x, a) \in \mathbb{K}_A. \tag{44}$$

Thus, by (39), (40), and Assumption 4(a) we obtain,

$$\hat{j}\tau(x, a, g) \leq c(x, a, g) + \int_{\mathbb{X}} \tau \hat{h}(y) Q(dy \mid x, a) - \tau \hat{h}(x) + \epsilon,$$

which implies

$$h^*(x) \leq c(x, a, g) - j^* \tau(x, a, g) + \int_{\mathbb{X}} h^*(y) Q(dy \mid x, a) + \epsilon, \tag{45}$$

with $j^* := \hat{j}$ and $h^*(\cdot) := \tau \hat{h}(\cdot)$. Since ϵ is arbitrary, we have

$$h^*(x) \leq \sup_{\theta \in \Theta(x,a)} \left\{ c(x, a, \theta) - j^* \tau(x, a, \theta) + \int_{\mathbb{X}} h^*(y) Q(dy \mid x, a) \right\} \quad \forall (x, a) \in \mathbb{K}_A,$$

and therefore,

$$h^*(x) \leq \inf_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} \left\{ c(x, a, \theta) - j^* \tau(x, a, \theta) + \int_{\mathbb{X}} h^*(y) Q(dy \mid x, a) \right\} \quad \forall x \in \mathbb{X}. \tag{46}$$

Now, from (43) there is $\hat{f} \in \mathbb{F}_{\mathbb{A}}$ such that for all $x \in \mathbb{X}$,

$$\hat{j} + \hat{h}(x) = \sup_{\theta \in \Theta(x, \hat{f})} \left\{ \hat{c}(x, \hat{f}, \theta) + \int_{\mathbb{X}} \hat{h}(y) \hat{Q}(dy \mid x, \hat{f}, \theta) \right\}.$$

Then, for all $x \in \mathbb{X}$ and $\theta \in \Theta(x, \hat{f})$,

$$\hat{j} + \hat{h}(x) \geq \hat{c}(x, \hat{f}, \theta) + \int_{\mathbb{X}} \hat{h}(y) \hat{Q}(dy \mid x, \hat{f}, \theta),$$

which yields (similarly as (45))

$$h^*(x) \geq c(x, \hat{f}, \theta) - j^* \tau(x, \hat{f}, \theta) + \int_{\mathbb{X}} h^*(y) Q(dy \mid x, \hat{f}).$$

Since θ is arbitrary in $\Theta(x, \hat{f})$, we have

$$h^*(x) \geq \sup_{\theta \in \Theta(x, \hat{f})} \left\{ c(x, \hat{f}, \theta) - j^* \tau(x, \hat{f}, \theta) + \int_{\mathbb{X}} h^*(y) Q(dy \mid x, \hat{f}) \right\},$$

and this implies that

$$h^*(x) \geq \inf_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} \left\{ c(x, a, \theta) - j^* \tau(x, a, \theta) + \int_{\mathbb{X}} h^*(y) Q(dy \mid x, a) \right\}. \tag{47}$$

Combining (46) and (47) we obtain

$$h^*(x) = \inf_{a\in A(x)} \sup_{\theta\in\Theta(x,a)} \left\{ c(x,a,\theta) - j^*\tau(x,a,\theta) + \int_{\mathbb{X}} h^*(y)Q(dy\mid x,a) \right\}, \quad x\in\mathbb{X}. \tag{48}$$

In addition, observe that the definitions of the function $h^* \in \mathbb{L}_W$ and the constant j^*, together Proposition 1 and Assumption 4(b) show the existence of $f^* \in \mathbb{F}_{\mathbb{A}}$ such that for all $x\in\mathbb{X}$

$$h^*(x) = \sup_{\theta\in\Theta(x,f^*)} \left\{ c(x,f^*,\theta) - j^*\tau(x,f^*,\theta) + \int_{\mathbb{X}} h^*(y)Q(dy\mid x,f^*) \right\}. \tag{49}$$

Finally, to conclude we prove that j^* is the optimal average cost and f^* is an average minimax strategy. From (49), for all $\theta\in\Theta(x,f^*)$, $x\in\mathbb{X}$,

$$h^*(x) \geq c(x,f^*,\theta) - j^*\tau(x,f^*,\theta) + \int_{\mathbb{X}} h^*(y)Q(dy\mid x,f^*). \tag{50}$$

Hence, if $\pi'\in\Pi_\Theta$ is an arbitrary strategy, iteration of (50) yields

$$h^*(x) \geq E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1} c(x_k,a_k,\theta_k)\right] - j^* E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1}\tau(x_k,a_k,\theta_k)\right] + E_x^{f^*\pi'}\left[h^*(x_n)\right], \quad n\in\mathbb{N}, \tag{51}$$

which in turn implies

$$j^* \geq \frac{E_x^{f^*\pi'}[\sum_{k=0}^{n-1} c(x_k,a_k,\theta_k)] + E_x^{f^*\pi'}[h^*(x_n)] - h^*(x)}{E_x^{f^*\pi'}[\sum_{k=0}^{n-1}\tau(x_k,a_k,\theta_k)]}, \tag{52}$$

where

$$E_x^{f^*\pi'}[h^*(x_n)] = \int_{\mathbb{X}} h^*(y)Q^n(dy\mid x,f^*).$$

Then, noting that (see Assumption 4(a) and (37))

$$nm \leq E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1}\tau(x_k,a_k,\theta_k)\right] \leq nM$$

and

$$\frac{E_x^{f^*\pi'}[h^*(x_n)]}{n} \to 0 \quad \text{as } n\to\infty,$$

and by taking lim sup as $n\to\infty$ in (52), it follows that (see (11))

$$j^* \geq J(x,f^*,\pi'). \tag{53}$$

Since π' is arbitrary we obtain

$$j^* \geq \sup_{\pi'\in\Pi_\Theta} J(x,f^*,\pi') \geq \inf_{\pi\in\Pi_{\mathbb{A}}} \sup_{\pi'\in\Pi_\Theta} J(x,\pi,\pi'). \tag{54}$$

On the other hand, from (48), for all $(x,a)\in\mathbb{K}_A$,

$$h^*(x) \leq \sup_{\theta\in\Theta(x,a)} \left\{ c(x,a,\theta) - j^*\tau(x,a,\theta) + \int_{\mathbb{X}} h^*(y)Q(dy\mid x,a) \right\}.$$

Then, (see (44)), for $\varepsilon > 0$ there is $g \in \mathbb{F}_\Theta$ such that

$$h^*(x) \le c(x,a,g(x,a)) - j^*\tau(x,a,g(x,a)) + \int_{\mathbb{X}} h^*(y)Q(dy \mid x,a) + \varepsilon.$$

If π is an arbitrary policy in $\Pi_{\mathbb{A}}$, then by applying similar arguments as (51)–(53), and using the fact (see (35))

$$E_x^{\pi g}\left[\sum_{k=0}^{n-1}\tau(x_k,a_k,\theta_k)\right] \ge nm, \quad n \in \mathbb{N},$$

we have

$$h^*(x) \le E_x^{\pi g}\left[\sum_{k=0}^{n-1}c(x_k,a_k,\theta_k)\right] - j^*E_x^{\pi g}\left[\sum_{k=0}^{n-1}\tau(x_k,a_k,\theta_k)\right] + E_x^{\pi g}\left[h^*(x_n)\right] + n\varepsilon,$$

which in turn yields

$$j^* \le J(x,\pi,g) \le \sup_{\pi' \in \Pi_\Theta} J(x,\pi,\pi').$$

Since π is arbitrary, we obtain

$$j^* \le \inf_{\pi \in \Pi_{\mathbb{A}}} \sup_{\pi' \in \Pi_\Theta} J(x,\pi,\pi'). \tag{55}$$

Finally, combining (54) and (55) we have

$$j^* = \sup_{\pi' \in \Pi_\Theta} J(x,f^*,\pi') = \inf_{\pi \in \Pi_{\mathbb{A}}} \sup_{\pi' \in \Pi_\Theta} J(x,\pi,\pi'),$$

and this completes the proof of Theorem 2.

7 Example: A Semi-Markov Inventory System

Consider an inventory-production system in which the demand of the product (immediately met), occurs at epochs $T_1, T_2, \ldots$ such that $\delta_n := T_n - T_{n-1}$ ($n = 1, 2, \ldots,$ $T_0 = 0$) is a random variable. The state variable x_n represents the stock level at the epoch T_n after the demand is met. The control or action variable a_n is the quantity ordered or produced in the beginning of the epoch T_n. The demand at epoch T_n is represented by ξ_n which are assumed to be no negatives i.i.d. random variables. According to this description, the evolution of the inventory process on the decision epochs T_n is given by the equation

$$x_{n+1} = (x_n + a_n - \xi_{n+1})^+, \quad n = 0, 1, 2, \ldots,$$

where x_0 is known.

To illustrate our assumptions, we fix the components of the semi-Markov control models (1) and (3) as follows. The state space as well as the controller and opponent actions spaces are $X = [0,\infty)$, $A = A(x) = [0,a^*]$, and $\Theta = \Theta(x,a) = [\theta_1,\theta_2] \subset \mathbb{R}$, for all $x \in X$, $a \in A$, and some $a^* > 0, \theta_1, \theta_2 \in \mathbb{R}$, respectively. We assume that the holding times have exponential distribution depending of an unknown parameter $\theta \in [\theta_1,\theta_2]$, as is stated in the following condition.

Assumption 6 There are positive numbers h_1, h_2 with $h_1 < h_2$ and a continuous function $h : \mathbb{K} \to [h_1, h_2]$ such that $F(t \mid x, a, \theta) = 0$ for $t < 0$ and

$$F(t \mid x, a, \theta) = 1 - e^{-h(x,a,\theta)t} \quad t \geq 0.$$

We impose the following condition on the demand.

Assumption 7 The random variable ξ_1 has a continuous bounded density g and

$$E(a^* - \xi_1) < 0.$$

Observe that the transition law Q takes the form

$$Q(B \mid x, a) = \int_{\mathbb{R}} 1_B[(x + a - s)^+] g(s) ds, \quad B \in \mathcal{B}([0, \infty)). \tag{56}$$

Furthermore, for $q \geq 0$ we define

$$\Psi(q) := E(\exp(q(a^* - \xi)).$$

Then, $\Psi(0) = 1$, and from Assumption 7, $\Psi'(0) = E(a^* - \xi) < 0$. Hence, there exists $q_0 > 0$ such that $\lambda_0 := \Psi(q_0) < 1$.

We define the weight function $W : [0, \infty) \to [1, \infty)$ as

$$W(x) := \exp(q_0 x), \quad x \in [0, \infty),$$

and we take the cost functions D and d satisfying the Assumption 2(a).

Now, we will proceed to verify the Assumptions 1, 2, 3, 4, and 5.

Under Assumption 6, the regularity condition imposed in Assumption 1 is satisfied. Indeed, fix $\eta > 0$ and let $\varepsilon = \exp(-h_2 \eta)$. Then

$$F(\eta \mid x, a, \theta) = 1 - e^{-h(x,a,\theta)\eta} \leq 1 - e^{-h_2 \eta} = 1 - \varepsilon.$$

Assumption 2(b), (c) follow from the continuity of g and the fact that

$$\int_0^\infty v(y) Q(dy \mid x, a) = v(0) P(x + a \leq \xi) + \int_0^{x+a} v(x + a - s) g(s) ds$$

and

$$\int_0^\infty W(y) Q(dy \mid x, a) = W(0) P(x + a \leq \xi) + \exp q_0(x + a) \int_0^{x+a} \exp(-q_0 s) g(s) ds. \tag{57}$$

Moreover, Assumption 2(d), (e) trivially hold since the multifunctions $x \longmapsto A(x)$ and $(x, a) \longmapsto \Theta(x, a)$ are constants.

To verify Assumption 3(a), we fix a discount factor $\alpha > 0$ such that $\alpha > 2h_2 - h_1$. Then from (57)

$$\begin{aligned} \int_0^\infty W(y) Q(dy \mid x, a) &\leq 1 + \exp(q_0 x) \int_0^\infty \exp(q_0(a - s)) g(s) ds \\ &\leq 1 + \lambda_0 \exp(q_0 x) \leq 2 \exp(q_0 x). \end{aligned} \tag{58}$$

On the other hand,

$$\beta_\alpha(x,a,\theta) := \int_0^\infty \exp(-\alpha s) h(x,a,\theta)\exp(-h(x,a,\theta)s)ds$$
$$= \frac{h(x,a,\theta)}{\alpha + h(x,a,\theta)} \le \frac{h_2}{\alpha + h_1},$$

which in turns implies

$$\rho_\alpha := \sup_{(x,a,\theta)\in\mathbb{K}} \beta_\alpha(x,a,\theta) < \frac{1}{2}. \tag{59}$$

Therefore, defining $b := 2$, the relations (58) and (59) yield Assumption 3(a). The continuity of $F(t \mid x,a,\theta)$ on $\mathbb{K}$ for each $t \in \mathbb{R}$, follows from the continuity of h on $\mathbb{K}$.

In addition, Assumption 4 holds because

$$\tau(x,a,\theta) = \int_0^\infty t F(dt \mid x,a,\theta) = \frac{1}{h(x,a,\theta)} \le \frac{1}{h_1} =: M.$$

To verify Assumption 5, define $\mu(\cdot) = \delta_0(\cdot)$, the Dirac measure concentrated at $x = 0$, $\phi(x,a) := P(x + a - \xi \le 0)$ and $\gamma = \lambda_0$. Then, ϕ is continuous on $\mathbb{K}$ and Assumption 5(a) follows from (56). To verify Assumption 5(b), note that from (57) and the fact that $W(0) = \mu(W)$, we obtain

$$\int_0^\infty W(y)Q(dy \mid x,a) = \phi(x,a)\mu(W) + \exp(q_0 x)\int_0^\infty \exp(q_0(a - q_0 s))g(s)ds$$
$$\le \exp(q_0 x)\int_0^\infty \exp(q_0(a^* - s))g(s)ds + \phi(x,a)\mu(W)$$
$$= \gamma W(x) + \phi(x,a)\mu(W).$$

Finally, for all $f \in \mathbb{F}$,

$$\int_{\mathbb{X}} \phi(x, f(x))\mu(dx) = \phi(0, f(0)) = P(f(0) - \xi \le 0) \ge P(a^* - \xi \le 0) > 0,$$

which proves Assumption 5(c).

References

1. Altman, E., Hordijk, A.: Zero-sum Markov games and worst-case optimal control of queueing systems. Queueing Syst. Theory Appl. **21**, 415–447 (1995)
2. Coraluppi, S.P., Marcus, S.I.: Mixed risk-neutral/minimax control of discrete-time finite state Markov decision process. IEEE Trans. Autom. Control **45**, 528–532 (2000)
3. Dynkin, E.B., Yushkevich, A.A.: Controlled Markov Processes. Springer, New York (1979)
4. Federgruen, A., Tijms, H.C.: The optimality equation in average cost denumerable state semi-Markov decision problems. Recurrence conditions and algorithms. J. Appl. Probab. **15**, 356–373 (1978)
5. Federgruen, A., Schweitzer, P.J., Tijms, H.C.: Denumerable undiscounted semi-Markov decision processes with unbounded rewards. Math. Oper. Res. **8**, 298–313 (1983)
6. González-Trejo, T.J., Hernández-Lerma, O., Hoyos-Reyes, L.F.: Minimax control of discrete-time stochastic systems. SIAM J. Control Optim. **41**, 1626–1659 (2003)
7. Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York (1996)

8. Hordijk, A., Passchier, O., Spieksma, F.M.: Optimal control against worst case admission policies: A multichained stochastic game. Math. Methods Oper. Res. **45**, 281–301 (1997)
9. Jagannathan, R.: A minimax ordering policy for the infinite stage dynamic inventory problem. Manag. Sci. **24**, 1138–1149 (1978)
10. Jaskiewicz, A.: An approximation approach to ergodic semi-Markov control processes. Math. Methods Oper. Res. **54**, 1–19 (2001)
11. Jaskiewicz, A.: A fixed point approach to solve the average cost optimality equation for semi-Markov decision processes with Feller transition probabilities. Commun. Stat., Theory Methods **36**, 2559–2575 (2007)
12. Kalyanasundaram, S., Chong, E.K.P., Shroff, N.B.: Markov decision processes with uncertain transition rates: sensitivity and max-min control. Asian J. Control **6**(2), 253–269 (2004)
13. Küenle, H.-U.: Stochastiche Spiele und Entscheidungsmodelle. B. G. Teubner, Leipzig (1986)
14. Küenle, H.-U.: On the optimality of (s, S)-strategies in a minimax inventory model with average cost criterion. Optimization **22**, 123–138 (1991)
15. Kurano, M.: Discrete-time Markovian decision processes with an unknown parameter-average return criterion. J. Oper. Res. Soc. Jpn. **15**, 67–76 (1972)
16. Kurano, M.: Minimax strategies for average cost stochastic games with an application to inventory models. J. Oper. Res. Soc. Jpn. **30**, 232–247 (1987)
17. Luque-Vásquez, F., Hernández-Lerma, O.: Semi-Markov models with average costs. Appl. Math. **26**, 315–331 (1999)
18. Luque-Vásquez, F., Minjárez-Sosa, J.A.: Semi-Markov control processes with unknown holding times distribution under a discounted criterion. Math. Methods Oper. Res. **61**, 455–468 (2005)
19. Luque-Vásquez, F., Minjárez-Sosa, J.A., Rosas-Rosas, L.C.: Semi-Markov control processes with unknown holding times distribution under an average cost criterion. Appl. Math. Optim. **61**, 217–336 (2010)
20. Mandl, P.: Estimation and control in Markov chains. Adv. Appl. Probab. **6**, 40–60 (1974)
21. Milliken, P., Marsh, C., Van Brunt, B.: Minimax controller design for a class of uncertain nonlinear systems. Automatica **35**, 583–590 (1999)
22. Puterman, M.L.: Markov Decision Processes. Discrete Stochastic Dynamic Programming. Wiley, New York (1994)
23. Rieder, U.: Measurable selection theorems for optimization problems. Manuscripta Math. 115–131 (1978)
24. Savkin, A.V., Peterson, I.R.: Minimax optimal control of uncertain systems with structured uncertainty. Int. J. Robust Nonlinear Control **5**, 119–137 (1995)
25. Schäl, M.: Conditions for optimality and for the limit on n-stage optimal policies to be optimal. Z. Wahrscheinlichkeitstheor. Verw. Geb. **32**, 179–196 (1975)
26. Schweitzer, P.J.: Iterative solution of the functional equations of undiscounted Markov renewal programming. J. Math. Anal. Appl. **34**, 495–501 (1971)
27. Vega-Amaya, O.: The average cost optimality equation: a fixed point approach. Bol. Soc. Mat. Mexicana **9**, 185–195 (2003)
28. Yu, W., Guo, X.: Minimax controller design for discrete-time time-varying stochastic systems. In: Proceedings of the 41st IEEE CDC, Las Vegas, Nevada, USA, pp. 598–603 (2002)