**Morfismos**, Vol. 6, No. 1, 2002, pp. 15-29

# Zero-sum semi-Markov games in Borel spaces with discounted payoff *

Fernando Luque-Vásquez [1]

**Abstract**

We study two-person zero-sum semi-Markov games in Borel spaces with possibly unbounded payoff, under the discounted criterion. We consider the $n$-stage case as well as the infinite horizon case. Conditions are given for the existence of the value of the game, the existence of optimal strategies for both players, and for a characterization of the optimal stationary policies.

*2000 Mathematics Subject Classification: 91A15, 91A25, 90C40.*
*Keywords and phrases: Zero-sum semi-Markov games, Borel spaces, discounted payoff, Shapley equation.*

## 1 Introduction

This paper deals with two-person zero-sum semi-Markov games with Borel spaces and possibly unbounded payoff function, under the discounted criterion. We consider the $n$-stage case as well as the infinite horizon case. Under suitable assumptions on the transition law, the payoff function and the distribution of the transition times, we show the existence of the value of the game, the existence of optimal strategies for both players, and we also obtain a characterization for a pair of stationary stategies to be optimal in the infinite horizon case.

Markovian stochastic games with discounted payoff have been studied by several authors (for example, [1, 7, 8, 10, 11, 12]) but, to the best of our knowledge, the only paper that studies semi-Markov stochastic

*Invited article.

[1]Partially supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT) Grant 28309E.

games with discounted payoff is [5], which considers a countable state space and a bounded payoff function. Our main results generalize to the semi-Markov context some theorems in [8, 10, 11] on which our approach is based. We also extend results in [5, 6].

The remainder of the paper is organized as follows. In Section 2 the semi-Markov game model is described. Next, in Section 3, the discounted criterion is introduced. In Section 4 we introduce the assumptions and present our main results, Theorems 4.3 and 4.4, which are proved in Sections 5 and 6, respectively.

**Terminology and notation**. Given a Borel space $X$, i.e. a Borel subset of a complete and separable metric space, we denote by $\mathcal{B}(X)$ its Borel $\sigma$-algebra. $\mathbb{P}(X)$ denotes the family of probability measures on $X$ endowed with the weak topology. If $X$ and $Y$ are Borel spaces, we denote by $\mathbb{P}(X\,|Y)$ the family of transition probabilities (or stochastic kernels) from $Y$ to $X$. For a transition probability $f \in \mathbb{P}(X\,|Y)$, we write its values as $f(y)(B)$ or $f(B\,|y)$ for all $B \in \mathcal{B}(X)$ and $y \in Y$. If $X = Y$, then $f$ is called a Markov transition probability on $X$.

## 2 The semi-Markov game

A semi-Markov game model is defined by a collection

$$(X, A, B, \mathbf{K}_A, \mathbf{K}_B, Q, F, , r),$$

where $X$ is the *state space*, and $A$ and $B$ are the *action spaces* for players 1 and 2, respectively. These spaces are assumed to be Borel spaces, whereas $\mathbf{K}_A \in \mathcal{B}(X \times A)$ and $\mathbf{K}_B \in \mathcal{B}(X \times B)$ are the *constraint sets*. For each $x \in X$, the $x$-section $A(x) := \{a \in A : (x, a) \in \mathbf{K}_A\}$ represents the set of admissible actions for player 1 in state $x$. Similarly, the $x$-section $B(x) := \{b \in B : (x, b) \in \mathbf{K}_B\}$ denotes the set of admissible actions for player 2 in state $x$. Let $\mathbf{K} := \{(x, a, b) : x \in X,\ a \in A(x),\ b \in B(x)\}$, which is a Borel subset of $X \times A \times B$ (see [9]). Moreover, $Q(\cdot \mid x, a, b)$ is a stochastic kernel on $X$ given $\mathbf{K}$ called the *transition law*, and $F(\cdot \mid x, a, b)$ is a probability distribution function on $\mathbb{R}_+ := [0, \infty)$ given $\mathbf{K}$ called the *transition time distribution*. Finally, $r$ is a real-valued measurable function on $\mathbf{K}$ that denotes the *payoff function*, and it represents the reward for player 1 and the cost function for player 2.

The game is played as follows: if $x$ is the state of the game at some decision (or transition) epoch, and the players independently choose actions $a \in A(x)$ and $b \in B(x)$, then the following happens: player 1 receives an immediate reward $r(x,a,b)$, player 2 incurres in a cost $r(x,a,b)$, and the system moves to a new state according to the probability measure $Q(\cdot \mid x,a,b)$. The time until the transition occurs is a random variable having the distribution function $F(\cdot \mid x,a,b)$.

Let $H_0 := X$ and $H_n := (\mathbf{K} \times \mathbb{R}_+) \times H_{n-1}$ for $n = 1, 2, ...$ . For each $n$, an element

$$h_n = (x_0, a_0, b_0, \delta_1, ..., x_{n-1}, a_{n-1}, b_{n-1}, \delta_n, x_n)$$

of $H_n$ represents the "history" of the game up to the $nth$ decision epoch. A *strategy* $\pi$ for player 1 is a sequence $\pi = \{\pi_n : n = 0, 1, ...\}$ of stochastic kernels $\pi_n \in \mathbb{P}(A \mid H_n)$ such that

$$\pi_n(A(x_n) \mid h_n) = 1 \;\; \forall h_n \in H_n.$$

We denote by $\Pi$ the family of all strategies for player 1. A strategy $\pi = \{\pi_n\}$ is called a *Markov strategy* if $\pi_n \in \mathbb{P}(A \mid X)$ for each $n = 0, 1, ...$ , that is, each $\pi_n$ depends only on the current state $x_n$ of the system. The set of all Markov strategies of player 1 is denoted by $\Pi_M$. Let $\Phi_1$ denote the class of all transition probabilities $f \in \mathbb{P}(A \mid X)$ such that $f(x) \in \mathbb{P}(A(x))$. A Markov strategy $\pi = \{\pi_n\}$ is said to be a *stationary strategy* if there exists $f \in \Phi_1$ such that $\pi_n = f$ for each $n = 0, 1, ...$ . In this case, the strategy is identified with $f$, and the set of all stationary strategies for player 1 with $\Phi_1$. The sets $\Gamma$, $\Gamma_M$ and $\Phi_2$ of all strategies, all Markov strategies and all stationary strategies for player 2 are defined similarly.

Let $(\Omega, \mathcal{F})$ be the canonical measurable space that consists of the sample space $\Omega := (X \times A \times B \times \mathbb{R}_+)^\infty$ and its product $\sigma$-algebra $\mathcal{F}$. Then for each pair of strategies $(\pi, \gamma) \in \Pi \times \Gamma$ and each initial state $x$ there exist a unique probability measure $P_x^{\pi\gamma}$ and a stochastic process $\{(x_n, a_n, b_n, \delta_{n+1}),\ n = 0, 1, ...\}$, where $x_n$, $a_n$ and $b_n$ represent the state and the actions for players 1 and 2, respectively, at the $nth$ decision epoch, whereas $\delta_n$ represents the time between the $(n-1)th$ and the $nth$ decision epoch. $E_x^{\pi\gamma}$ denotes the expectation operator with respect to $P_x^{\pi\gamma}$.

## 3 Optimality criteria

We assume that rewards and costs are continuously discounted and player 1 tries to maximize the expected discounted payoff, while player

2 tries to minimize it.

**Definition 3.1**. For $n \geq 1$, $\alpha > 0$, $x \in X$ and $(\pi, \gamma) \in \Pi \times \Gamma$, the *expected $n$-stage $\alpha$-discounted payoff* is defined as

$$V_n(x, \pi, \gamma) := E_x^{\pi\gamma} \sum_{k=0}^{n-1} e^{-\alpha T_k} r(x_k, a_k, b_k), \tag{1}$$

where $T_0 = 0$ and $T_n = T_{n-1} + \delta_n$. The *infinite-horizon total expected $\alpha$-discounted payoff* is

$$V(x, \pi, \gamma) := E_x^{\pi\gamma} \sum_{k=0}^{\infty} e^{-\alpha T_k} r(x_k, a_k, b_k). \tag{2}$$

To define our optimality criteria, we need to introduce the following concepts. The functions on $X$ given by

$$L(x) := \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} V(x, \pi, \gamma) \quad \text{and} \quad U(x) := \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} V(x, \pi, \gamma) \tag{3}$$

are called the *lower value* and the *upper value*, respectively, of the (expected) $\alpha$-discounted payoff game. It is clear that $L(\cdot) \leq U(\cdot)$ in general, but if it holds that $L(x) = U(x)$ for all $x \in X$, then the common value is called the *value* of the semi-Markov game and will be denoted by $V^*(x)$.

In Section 3 we give assumptions that guarantee that the functions in (1), (2) and (3) are well defined.

**Definition 3.2**. (a) A strategy $\pi^* \in \Pi$ is said to be *$\alpha$-optimal for player* 1 if

$$U(x) \leq V(x, \pi^*, \gamma) \quad \forall \gamma \in \Gamma,\ x \in X.$$

(b) A strategy $\gamma^* \in \Gamma$ is said to be *$\alpha$-optimal for player* 2 if

$$V(x, \pi, \gamma^*) \leq L(x) \quad \forall \pi \in \Pi,\ x \in X.$$

(c) A pair $(\pi^*, \gamma^*) \in \Pi \times \Gamma$ is said to be an *$\alpha$-optimal strategy pair* if, for all $x \in X$,

$$U(x) = \inf_{\gamma \in \Gamma} V(x, \pi^*, \gamma) \quad \text{and} \quad L(x) = \sup_{\pi \in \Pi} V(x, \pi, \gamma^*).$$

We note that the existence of an $\alpha$-optimal strategy either for player 1 or player 2, implies that the game has a value.

For the $n$-stage semi-Markov game, the lower value $L_n$, the upper value $U_n$, the value $V_n^*$ and optimal strategies are defined similarly.

**Remark 3.3.** Let

$$\beta_\alpha(x,a,b) := \int_0^\infty e^{-\alpha t} F(dt \mid x,a,b). \tag{4}$$

Then using properties of the conditional expectation we can write

$$V(x,\pi,\gamma) = E_x^{\pi\gamma}[r(x_0,a_0,b_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \beta_\alpha(x_k,a_k,b_k) r(x_n,a_n,b_n)], \tag{5}$$

and for $n \geq 1$

$$V_n(x,\pi,\gamma) = E_x^{\pi\gamma}[r(x_0,a_0,b_0) + \sum_{k=1}^{n-1} \prod_{i=0}^{k-1} \beta_\alpha(x_i,a_i,b_i) r(x_k,a_k,b_k)].$$

## 4 Assumptions and main results

The problem we are concerned with is to show the existence of $\alpha$-optimal strategies which, as is well known (see for instance [7]), requires imposing suitable assumptions on the semi-Markov game model. The first one is a regularity condition that ensures that an infinite number of transitions do not occur in a finite interval. The second one is a combination of standard continuity and compactness requirements, whereas the third one is a growth condition on the payoff function $r$.

**Assumption 1** (A1). There exist $\theta > 0$ and $\varepsilon > 0$ such that

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

An important consequence of this assumption is the following.

**Lemma 4.1**. If A1 holds, then

$$\rho_\alpha := \sup_{(x,a,b)\in\mathbf{K}} \beta_\alpha(x,a,b) < 1. \tag{6}$$

*Proof:* Let $(x, a, b) \in \mathbf{K}$ be fixed. Then integrating by parts in (4) we have

$$\begin{aligned}\beta_\alpha(x, a, b) &= \alpha \int_0^\infty e^{-\alpha t} F(t\,|x, a, b)dt \\ &= \alpha[\int_0^\theta e^{-\alpha t} F(t\,|x, a, b)dt + \int_\theta^\infty e^{-\alpha t} F(t\,|x, a, b)dt] \\ &\le (1-\varepsilon)(1-e^{-\alpha\theta}) + e^{-\alpha\theta} = 1 - \varepsilon + \varepsilon e^{-\alpha\theta} < 1.\end{aligned}$$

As $(x, a, b) \in \mathbf{K}$ was arbitrary, we get (6). □

**Assumption 2** (A2). (a) For each $x \in X$, the sets $A(x)$ and $B(x)$ are compact.

(b) For each $(x, a, b) \in \mathbf{K}$, $r(x, \cdot, b)$ is upper semicontinuous on $A(x)$, and $r(x, a, \cdot)$ is lower semicontinuous on $B(x)$.

(c) For each $(x, a, b) \in \mathbf{K}$ and each bounded and measurable function $v$ on $X$, the functions

$$a \longmapsto \int v(y)Q(dy \mid x, a, b) \quad \text{and} \quad b \longmapsto \int v(y)Q(dy \mid x, a, b)$$

are continuous on $A(x)$ and $B(x)$, respectively.

(d) For each $t \ge 0$, $F(t \mid x, a, b)$ is continuous on $\mathbf{K}$.

**Assumption 3** (A3). There exist a measurable function $w : X \to [1, \infty)$ and positive constants $m$ and $\eta$, with $\eta\rho_\alpha < 1$, such that for all $(x, a, b) \in \mathbf{K}$

(a) $|r(x, a, b)| \le mw(x)$;

(b) $\int w(y)Q(dy \mid x, a, b) \le \eta w(x)$.

In addition, part (c) in A2 holds when $v$ is replaced with $w$.

**Remark 4.2.** By Lemma 1.11 in [8], it follows that if Assumption 2(a) holds then the multifunctions $\mathbb{A} : X \to 2^{\mathbb{P}(A)}$ and $\mathbb{B} : X \to 2^{\mathbb{P}(B)}$ defined as $\mathbb{A}(x) := \mathbb{P}(A(x))$ and $\mathbb{B}(x) := \mathbb{P}(B(x))$ are measurable compact-valued multifunctions.

We now introduce the following notation: for any given function $h : \mathbf{K} \to \mathbb{R}$, $x \in X$, and probability measures $\mu \in \mathbb{A}(x)$ and $\lambda \in \mathbb{B}(x)$ we write

$$h(x, \mu, \lambda) := \int_{B(x)} \int_{A(x)} h(x, a, b)\mu(da)\lambda(db),$$

whenever the integrals are well defined. In particular,

$$r(x,\mu,\lambda) := \int_{B(x)} \int_{A(x)} r(x,a,b)\mu(da)\lambda(db),$$

$$\beta_\alpha(x,\mu,\lambda) := \int_{B(x)} \int_{A(x)} \beta_\alpha(x,a,b)\mu(da)\lambda(db),$$

and

$$Q(D \mid x,\mu,\lambda) := \int_{B(x)} \int_{A(x)} Q(D \mid x,a,b)\mu(da)\lambda(db), \quad D \in \mathcal{B}(X).$$

$B_w(X)$ denotes the linear space of measurable functions $u$ on $X$ with finite $w$-norm, which is defined as

$$\|u\|_w := \sup_{x\in X} \frac{|u(x)|}{w(x)}.$$

For $u \in B_w(X)$ and $(x,a,b) \in \mathbf{K}$, we write

$$H(u,x,a,b) := r(x,a,b) + \beta_\alpha(x,a,b) \int_X u(y)Q(dy \mid x,a,b).$$

For each function $u \in B_w(X)$ let

$$T_\alpha u(x) := \sup_{\mu\in\mathbb{A}(x)} \inf_{\lambda\in\mathbb{B}(x)} H(u,x,\mu,\lambda), \tag{7}$$

which defines a function $T_\alpha u$ in $B_w(X)$ (see Lemma 5.1 below). We call $T_\alpha$ the Shapley operator, and a function $v \in B_w(X)$ is said to be a solution to the *Shapley equation* if $T_\alpha v(x) = v(x)$ for all $x \in X$. In the proof of Lemma 5.1, we show that if Assumptions 1, 2 and 3 hold, then for $\mu \in \mathbb{A}(x)$, $H(u,x,\mu,\cdot)$ is l.s.c. on $\mathbb{B}(x)$, and for $\lambda \in \mathbb{B}(x)$, $H(u,x,\cdot,\lambda)$ is u.s.c. on $\mathbb{A}(x)$. Thus, by Theorem A.2.3 in [2] the supremum and the infimum are indeed attained in (7). Hence, we can write

$$T_\alpha u(x) := \max_{\mu\in\mathbb{A}(x)} \min_{\lambda\in\mathbb{B}(x)} H(u,x,\mu,\lambda).$$

We are now ready to state our main results.

**Theorem 4.3.** Suppose that A1-A3 hold. Then the $n$-stage semi-Markov game ($n \geq 1$) has a value $V_n^* \in B_w(X)$ and both players have $\alpha$-optimal Markov strategies. Moreover, for each $n \geq 2$,

$$V_n^*(x) = T_\alpha V_{n-1}^*(x).$$

**Theorem 4.4.** If A1-A3 hold, then

(a) The semi-Markov game has a value $V^*$, which is the unique function in $B_w(X)$ that satisfies the Shapley equation,

$$V^*(x) = T_\alpha V^*(x),$$

and, furthermore, there exists an $\alpha$-optimal strategy pair.

(b) A pair of stationary strategies $(f, g) \in \Phi_1 \times \Phi_2$ is $\alpha$-optimal if and only if $V(\cdot, f, g)$ is a solution to the Shapley equation.

# 5 Proof of Theorem 4.3

First we shall prove a preliminary result.

**Lemma 5.1.** If A1-A3 hold, then for each $u \in B_w(X)$, the function $T_\alpha u$ is in $B_w(X)$, and

$$T_\alpha u(x) = \min_{\lambda \in \mathbb{B}(x)} \max_{\mu \in \mathbb{A}(x)} H(u, x, \mu, \lambda). \tag{8}$$

Moreover, there exist stationary strategies $f \in \Phi_1$ and $g \in \Phi_2$ such that

$$\begin{aligned} T_\alpha u(x) &= H(u, x, f(x), g(x)) \\ &= \max\nolimits_{\mu \in \mathbb{A}(x)} H(u, x, \mu, g(x)) \\ &= \min\nolimits_{\lambda \in \mathbb{B}(x)} H(u, x, f(x), \lambda). \end{aligned} \tag{9}$$

*Proof:* By Lemma 4.1 and A3, we have that for $u \in B_w(X)$ and $(x, a, b) \in \mathbf{K}$,

$$|H(u, x, a, b)| \leq mw(x) + \rho_\alpha \|u\|_w \eta w(x),$$

which, as $T_\alpha u$ is measurable, implies that $T_\alpha u \in B_w(X)$. On the other hand, by A2, it follows that the function $x \longmapsto H(u, x, a, b)$ is in $B_w(X)$ and $H(u, x, \cdot, b)$ is u.s.c. on $A(x)$. Then, for fixed $\lambda \in \mathbb{B}(x)$, by Fatou's Lemma, the function

$$a \longmapsto \int_{B(x)} H(u, x, a, b) \lambda(db)$$

is u.s.c. and bounded on $A(x)$. Thus, since convergence on $\mathbb{A}(x)$ is the weak convergence of probability measures, by Theorem 2.8.1 in [2], the function $H(u,x,\cdot,\lambda)$ is u.s.c. on $\mathbb{A}(x)$. Similarily, $H(u,x,\mu,\cdot)$ is l.s.c. on $\mathbb{B}(x)$. Moreover, $H(u,x,\mu,\lambda)$ is concave in $\mu$ and convex in $\lambda$. Thus, by Fan's minimax Theorem [3] we obtain (8). The existence of stationary strategies $f\in\Phi_1$ and $g\in\Phi_2$ that satisfy (9) follows from (8) and well-known measurable selection theorems (see for instance Lemma 4.3 in [8]). $\square$

**Proof of Theorem 4.3.** The proof proceeds by induction. For $n=1$, the theorem follows directly from Definition 3.1 and Lemma 5.1 with $u(\cdot)=0$. Suppose the result holds for $n-1$ ($n\geq 2$). Let $\pi^{(n-1)}=(f_1,f_2,...,f_{n-1})$ with $f_i\in\Phi_1$ and $\gamma^{(n-1)}=(g_1,g_2,...,g_{n-1})$ with $g_i\in\Phi_2$ be a pair of $\alpha$-optimal Markov strategies of players 1 and 2, respectively, in the $(n-1)$-stage semi-Markov game. Then

$$V^*_{n-1}(\cdot)=V_{n-1}(\cdot,\pi^{(n-1)},\gamma^{(n-1)}), \tag{10}$$

and

$$V^*_{n-1}(\cdot)=T_\alpha V^*_{n-2}(\cdot).$$

For an arbitrary $g\in\Phi_2$ put $\gamma^g=(g,g_1,...,g_{n-1})$. By definition of $U_n$, we have

$$U_n(x)\leq \sup_{\pi\in\Pi} V_n(x,\pi,\gamma^g),$$

from which we obtain

$$\begin{aligned}U_n(x)\leq \sup_{\mu\in\mathbb{A}(x)}\{&\int_{B(x)}\int_{A(x)}[r(x,a,b)\\ &+\beta_\alpha(x,a,b)\int \sup_{\pi\in\Pi} V_{n-1}(y,\pi,\gamma^{(n-1)})Q(dy\,|x,a,b)]\mu(da)g(x)(db)\}.\end{aligned}$$

Therefore, by the induction hypothesis,

$$U_n(x)\leq \sup_{\mu\in\mathbb{A}(x)} H(V^*_{n-1},x,\mu,g(x)).$$

Hence, since $g\in\Phi_2$ was arbitrary,

$$U_n(x)\leq \inf_{\lambda\in\mathbb{B}(x)}\sup_{\mu\in\mathbb{A}(x)} H(V^*_{n-1},x,\mu,\lambda),$$

and, by Lemma 5.1,

$$U_n(x)\leq T_\alpha V^*_{n-1}(x). \tag{11}$$

Similarily we obtain

$$L_n(x) \geq T_\alpha V_{n-1}^*(x). \tag{12}$$

Combining (11) and (12) we get $L_n(x) = U_n(x) = T_\alpha V_{n-1}^*(x)$, i.e. the $n$-stage semi-Markov game has a value $V_n^*$ and $V_n^* = T_\alpha V_{n-1}^*$. Further, by Lemma 5.1 $V_n^* \in B_w(X)$, and there exist $f_0 \in \Phi_1$ and $g_0 \in \Phi_2$ such that for every $f \in \Phi_1$ and $g \in \Phi_2$,

$$\begin{aligned} H(V_{n-1}^*, x, f(x), g_0(x)) &\leq V_n^*(x) \\ &= H(V_{n-1}^*, x, f_0(x), g_0(x)) \\ &\leq H(V_{n-1}^*, x, f_0(x), g(x)). \end{aligned} \tag{13}$$

Let $\pi^{(n)} = (f_0, f_1, ..., f_{n-1})$ and $\gamma^{(n)} = (g_0, g_1, ..., g_{n-1})$. Then, from (10) and (13) it follows that $\pi^{(n)}$ and $\gamma^{(n)}$ are $\alpha$-optimal strategies for players 1 and 2, respectively. □

# 6 Proof of Theorem 4.4

To prove Theorem 4.4, we need some preliminary lemmas for which we require the following notation. For a pair of stationary strategies $(f, g) \in \Phi_1 \times \Phi_2$, we define the operator $T_{fg}$ on $B_w(X)$ as:

$$T_{fg}u(x) := H(u, x, f(x), g(x)).$$

It is clear (see the proof of Lemma 5.1) that $T_{fg}u$ belongs to $B_w(X)$ for each $u \in B_w(X)$.

**Lemma 6.1.** If A1-A3 hold, then both $T_\alpha$ and $T_{fg}$ are contraction operators with modulus $\eta\rho_\alpha < 1$.

*Proof:* First we note that both operators are monotone. That is, if $u, v \in B_w(X)$ and $u(\cdot) \leq v(\cdot)$, then for all $x \in X$

$$T_{fg}u(x) \leq T_{fg}v(x).$$

Similarly, $T_\alpha u(x) \leq T_\alpha v(x)$ for all $x \in X$. Also, it is easy to see that for $k \geq 0$,

$$T_{fg}(u + kw)(x) \leq T_{fg}u(x) + \rho_\alpha \eta k w(x) \quad \forall x \in X,$$

and

$$T_\alpha(u + kw)(x) \leq T_\alpha u(x) + \rho_\alpha \eta k w(x) \quad \forall x \in X. \tag{14}$$

Now, for $u, v \in B_w(X)$, by (14), the monotonocity of $T_\alpha$ and the fact that $u \leq v + w \|u - v\|_w$, it follows that

$$T_\alpha u(x) \leq T_\alpha v(x) + \rho_\alpha \eta \|u - v\|_w w(x) \ \ \forall x \in X,$$

so that

$$T_\alpha u(x) - T_\alpha v(x) \leq \rho_\alpha \eta \|u - v\|_w w(x) \ \ \forall x \in X. \tag{15}$$

If we now interchange $u$ and $v$ we obtain

$$T_\alpha u(x) - T_\alpha v(x) \geq -\rho_\alpha \eta \|u - v\|_w w(x) \ \ \forall x \in X, \tag{16}$$

and combining (15) and (16) we get

$$|T_\alpha u(x) - T_\alpha v(x)| \leq \rho_\alpha \eta \|u - v\|_w w(x) \ \ \forall x \in X,$$

i.e.

$$\|T_\alpha u - T_\alpha v\|_w \leq \rho_\alpha \eta \|u - v\|_w .$$

Hence, $T_\alpha$ is a contraction operator with modulus $\rho_\alpha \eta$. Using the same arguments we can prove that $T_{fg}$ is a contraction operator with the same modulus $\rho_\alpha \eta$. □

**Remark 6.2.** Since $T_\alpha$ and $T_{fg}$ are contraction operators, by Banach's Fixed Point Theorem there exist functions $v^*$ and $v_{fg}$ in $B_w(X)$ such that $T_\alpha v^*(x) = v^*(x)$ and $T_{fg} v_{fg}(x) = v_{fg}(x)$ for all $x \in X$.

**Lemma 6.3.** For a pair of stationary strategies $(f, g) \in \Phi_1 \times \Phi_2$, the function $V(\cdot, f, g)$ is the unique fixed point of $T_{fg}$ in $B_w(X)$.

*Proof:* We have to show that $V(x, f, g) = T_{fg} V(x, f, g) \ \forall x \in X$. Now,

$$\begin{aligned}
V(x, f, g) &= E_x^{fg}\{r(x_0, a_0, b_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) r(x_n, a_n, b_n)\} \\
&= r(x, f(x), g(x)) + E_x^{fg}\{\sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) r(x_n, a_n, b_n)\} \\
&= r(x, f(x), g(x)) + E_x^{fg}\{\beta_\alpha(x_0, a_0, b_0) E_x^{fg}[r(x_1, a_1, b_1) \\
&+ \sum_{n=2}^{\infty} \prod_{k=1}^{n-1} \beta_\alpha(x_k, a_k, b_k) r(x_n, a_n, b_n) \, |h_1]\} \\
&= r(x, f(x), g(x)) + E_x^{fg}\{\beta_\alpha(x_0, a_0, b_0) V(x_1, f, g)\} \\
&= T_{fg} V(x, f, g).
\end{aligned}$$

Thus, $V(\cdot, f, g)$ is the fixed point of $T_{fg}$. □

**Lemma 6.4.** Suppose that A1-A3 hold, and let $\pi$ and $\gamma$ be arbitrary strategies for players 1 and 2, respectively. Then for each $x \in X$ and $n = 0, 1, \ldots$

(a) $E_x^{\pi\gamma} w(x_n) \leq \eta^n w(x)$,

(b) $|E_x^{\pi\gamma} r(x_n, a_n, b_n)| \leq m\eta^n w(x)$,

(c) $\lim_{n\to\infty} E_x^{\pi\gamma}(\prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) u(x_n)) = 0$ for each $u \in B_w(X)$.

*Proof:* For $n = 0$, (a) and (b) are trivially satisfied. Now, if $n \geq 1$ then, by A3(b),

$$\begin{aligned} E_x^{\pi\gamma}[w(x_n) \mid h_{n-1}, a_{n-1}, b_{n-1}] &= \int w(y) Q(dy \mid x_{n-1}, a_{n-1}, b_{n-1}) \\ &\leq \eta w(x_{n-1}). \end{aligned}$$

Hence $E_x^{\pi\gamma} w(x_n) \leq \eta E_x^{\pi\gamma} w(x_{n-1})$, which by iteration yields (a). Part (b) follows immediately from (a) and A3(a). To prove (c), we observe that Lemma 4.1 and (a) yield

$$\begin{aligned} \left| E_x^{\pi\gamma}[\prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) u(x_n)] \right| &\leq \rho_\alpha^n E_x^{\pi\gamma} |u(x_n)| \leq \rho_\alpha^n \|u\|_w E_x^{\pi\gamma} w(x_n) \\ &\leq (\rho_\alpha \eta)^n \|u\|_w w(x). \end{aligned}$$

This yields (c), since $\rho_\alpha \eta < 1$. □

**Proof of Theorem 4.4.** (a) Let $V_\alpha$ be the unique fixed point in $B_w(X)$ of $T_\alpha$. Then

$$V_\alpha(x) = T_\alpha V_\alpha(x) = \max_{\mu\in\mathbb{A}(x)} \min_{\lambda\in\mathbb{B}(x)} H(V_\alpha, x, \mu, \lambda).$$

By Lemma 5.1 there exists a pair of stationary strategies $(f^*, g^*) \in \Phi_1 \times \Phi_2$ such that

$$\begin{aligned} V_\alpha(x) &= H(V_\alpha, x, f^*(x), g^*(x)) \\ &= \min_{\lambda\in\mathbb{B}(x)} H(V_\alpha, x, f^*(x), \lambda) \\ &= \max_{\mu\in\mathbb{A}(x)} H(V_\alpha, x, \mu, g^*(x)). \end{aligned} \tag{17}$$

We will prove that $V_\alpha$ is the value of the semi-Markov game and that $(f^*, g^*)$ is an $\alpha$-optimal strategy pair. The first equality in (17) implies that $V_\alpha$ is the fixed point in $B_w(X)$ of $T_{f^*g^*}$. Thus, by Lemma 6.3, $V_\alpha(\cdot) = V(\cdot, f^*, g^*)$, so that it is enough to show that for arbitrary $\pi \in \Pi$ and $\gamma \in \Gamma$,

$$V(x, f^*, \gamma) \geq V(x, f^*, g^*) \geq V(x, \pi, g^*) \quad \forall x \in X. \tag{18}$$

We will prove the second inequality in (18). A similar proof can be given for the first inequality. By (5) we have

$$V(x, \pi, g^*) = E_x^{\pi g^*}\{r(x_0, a_0, b_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) r(x_n, a_n, b_n)\}.$$

From properties of the conditional expectation we have for $n \geq 1$, $h_n \in H_n$, $a_n \in A(x_n)$, and $b_n \in B(x_n)$,

$$\begin{aligned}
&E_x^{\pi g^*}\{\textstyle\prod_{k=0}^{n} \beta_\alpha(x_k, a_k, b_k) V(x_{n+1}, f^*, g^*) \mid h_n, a_n, b_n\}\\
&= \textstyle\prod_{k=0}^{n} \beta_\alpha(x_k, a_k, b_k) E_x^{\pi g^*}\{V(x_{n+1}, f^*, g^*) \mid h_n, a_n, b_n\}\\
&= \textstyle\prod_{k=0}^{n} \beta_\alpha(x_k, a_k, b_k) \int V(y, f^*, g^*) Q(dy \mid x_n, \pi_n(h_n), g^*(x_n))\\
&= \textstyle\prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k)\{\beta_\alpha(x_n, a_n, b_n) \int V(y, f^*, g^*) Q(dy \mid x_n, \pi_n(h_n),\\
&\quad g^*(x_n)) \;+\; r(x_n, \pi_n(h_n), g^*(x_n)) - r(x_n, \pi_n(h_n), g^*(x_n))\}\\
&\leq \textstyle\prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k)[V(x_n, f^*, g^*) - r(x_n, \pi_n(h_n), g^*(x_n))].
\end{aligned}$$

Equivalently, for $n \geq 1$

$$\begin{aligned}
&\textstyle\prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) V(x_n, f^*, g^*)\\
&\quad - E_x^{\pi g^*}\{\textstyle\prod_{k=0}^{n} \beta_\alpha(x_k, a_k, b_k) V(x_{n+1}, f^*, g^*) \mid h_n, a_n, b_n]\\
&\quad \geq \textstyle\prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) r(x_n, \pi_n(h_n), g^*(x_n)).
\end{aligned}$$

We also have

$$V(x_0, f^*, g^*) - E_x^{\pi g^*}[\beta_\alpha(x_0, a_0, b_0) V(x_1, f^*, g^*)] \geq r(x_0, \pi_0(x_0), g^*(x_0)).$$

Now, taking expectations and summing over $n = 0, 1, ..., N$ we obtain

$$V(x, f^*, g^*) - E_x^{\pi g^*}\{\prod_{k=0}^{N} \beta_\alpha(x_k, a_k, b_k) V(x_{N+1}, f^*, g^*)\}$$
$$\geq E_x^{\pi g^*}[r(x_0, a_0, b_0) + \sum_{n=1}^{N} \prod_{k=0}^{n-1} \beta_\alpha(x_k, a_k, b_k) r(x_n, a_n, b_n)].$$

Finally, letting $N \to \infty$, by (5) and Lemma 6.4 (c) we obtain the required result.

(b) ($\Longrightarrow$) Suppose that $(f, g) \in \Phi_1 \times \Phi_2$ is a pair of $\alpha$-optimal stationary strategies. Then for all $x \in X$, $\pi \in \Pi$ and $\gamma \in \Gamma$,

$$V(x, f, \gamma) \geq V(x, f, g) \geq V(x, \pi, g). \tag{19}$$

Fix $x \in X$ and for an arbitrary $\lambda \in \mathbb{B}(x)$ define $\hat{\gamma} = (\hat{\gamma}_n)$ as follows: $\hat{\gamma}_0 = \lambda$ and $\hat{\gamma}_n = g$ for $n = 1, 2, ....$ Then, by the first inequality in (19),

$$V(x, f, g) \leq V(x, f, \hat{\gamma}) = \int_{B(x)} \int_{A(x)} [r(x, a, b)$$
$$+ \beta_\alpha(x, a, b) \int V(y, f, g) Q(dy \,|x, a, b)] f(x)(da) \lambda(db).$$

It follows that

$$V(x, f, g) \leq H(V(\cdot, f, g), x, f, \lambda),$$

from which we get

$$V(x, f, g) \leq T_\alpha V(x, f, g).$$

Similarily, we can prove

$$V(x, f, g) \geq T_\alpha V(x, f, g),$$

and combining the last two inequalities we get the desired result.

($\Longleftarrow$) The proof of this part is contained in the proof of part (a). □

**Acknowledgement**

The author wishes to thank Professor Onésimo Hernández-Lerma for his valuable comments and suggestions.

Fernando Luque-Vásquez
Departmento de Matemáticas,
Universidad de Sonora,
Rosales y Blvd. Luis Encinas,
Hermosillo, Sonora, MEXICO.
fluque@gauss.mat.uson.mx

## References

[1] Altman, E.; Hordijk, A.; Spieksma, F. M., *Contraction conditions for average and $\alpha$-discounted optimality in countable state Markov games with unbounded rewards*, Math. Oper. Res. **22** (1997), 588-618.

[2] Ash, R. B.; Doléans-Dade, C. A., Probability and Measure Theory, Academic Press, New York, 2000.

[3] Fan, K., *Minimax theorems*, Proc. Nat. Acad. Sci. USA **39** (1953), 42-47.

[4] Hernández-Lerma, O.; Lasserre, J. B., *Zero-sum stochastic games in Borel spaces: average payoff criteria*, SIAM J. Control Optim. **39** (2001), 1520-1539.

[5] Lal, A. K.; Sinha, S., *Zero-sum two person semi-Markov games*, J. Appl. Prob. **29** (1992), 56-72.

[6] Luque-Vásquez, L.; Robles-Alcaraz, M. T., *Controlled semi-Markov models with discounted unbounded costs*, Bol. Soc. Mat. Mex. **39** (1994), 51-68.

[7] Maitra, A.; Parthasarathy, T., *On stochastic games*, J. Optim. Theory Appl. **5** (1970), 289-300.

[8] Nowak, A. S., *On zero-sum stochastic games with general state space I*, Probab. Math. Statist. **4** (1984), 13-32.

[9] Nowak, A. S., *Measurable selection theorems for minimax stochastic optimization problems*, SIAM J. Control Optim. **23** (1985), 466-476.

[10] Nowak, A. S., *Optimal strategies in a class of zero-sum ergodic stochastic games*, Math. Meth. Oper. Res. **50** (1999), 399-419.

[11] Ramírez-Reyes, F., *Existence of optimal strategies for zero-sum stochastic games with discounted payoff*, Morfismos **5** (2001), 63-83.

[12] Sennott, L. I., *Zero-sum stochastic games with unbounded costs: discounted and average cost cases*, Z. Oper. Res. **39** (1994), 209-225.