Morfismos, Vol. 5, No. 1, 2001, pp. 63–83

Existence of optimal strategies for zero-sum stochastic games with discounted payoff [*]

Francisco Ramírez-Reyes

Abstract

This work deals with zero-sum stochastic games with Borel state and action spaces, and unbounded payoff function. We consider finite-horizon as well as infinite-horizon discounted problems. Our main purpose is to give conditions for the existence of the game value and for the existence of optimal strategies for both players. The infinite-horizon case is analyzed using the approach of successive approximations.

2000 Mathematics Subject Classification: 91A15, 91A25.
Keywords and phrases: Zero-sum stochastic games, Borel spaces, expected payoff function.

1 Introduction

We consider in this work two-person zero-sum stochastic games with Borel state and action spaces. Our main objective is to give conditions for the existence of optimal strategies for the discounted payoff criterion with unbounded running (or immediate) payoff function. We first study the finite-horizon problem in n-stages, and then obtain the infinite-horizon discounted case by letting $n \to \infty$.

Zero-sum stochastic games with discounted payoff have been widely studied for finite or countable state spaces and bounded payoff function.

[*] Research partially supported by a CONACyT scholarship. This paper is part of the author's M.Sc. Thesis presented at the Department of Mathematics of CINVESTAV-IPN.

The countable state case with unbounded rewards has been studied by Altman [1] and Sennott [21], for instance. For the uncountable case and bounded reward function, see [12,14,17] and their references, and for unbounded reward functions, see for example Rieder [20] and Nowak [16]. Nowak [16] has established the existence of optimal strategies under hypotheses similar to ours. However, our proof of the existence of optimal strategies is different from Nowak's.

For applications of stochastic games to queueing systems and telecommunications networks, see [1,21]. On the other hand, for results on the *average payoff* case, see for example, [5,8,13,16].

This work is organized as follows. Section 2 introduces the basic notation and definitions. In Section 3 we describe the game model and the optimality criteria we are interested in, as well as our assumptions. Section 4 deals with finite-horizon stochastic games. In that section we prove the existence of the game value, and the existence of optimal strategies for both players. Finally, these results are extended in Section 5 to the infinite-horizon stochastic game.

2 Preliminaries

In this section we introduce some basic notation and definitions used throughout the paper.

2.1 Borel spaces and stochastic kernels

A Borel subset X of a complete and separable metric space is called a *Borel space*, and its Borel σ-algebra is denoted by $\mathcal{B}(X)$. As we only consider Borel spaces, throughout the following, "measurable" (for either sets or functions) means "Borel measurable".

Definition 2.1.1 Let X and Y be Borel spaces. A *stochastic kernel* on X given Y is a function $P(\cdot|\cdot)$ such that

i) $P(\cdot|y)$ is a probability measure on X for each fixed $y \in Y$, and

ii) $P(D|\cdot)$ is a measurable function on Y for each fixed $D \in \mathcal{B}(X)$.

The set of all stochastic kernels on X given Y is denoted by $\mathbb{P}(X|Y)$.

We denote by $M(X)$ the set of all measurable functions $u : X \to \mathbb{R}$ and by $B(X)$ the subset of bounded functions in $M(X)$. By $C(X)$ we denote the set of all continuous function in $B(X)$. Thus, we have

$$C(X) \subset B(X) \subset M(X).$$

Given a Borel space X, we denote by $\mathbb{P}(X)$ the family of probability measures on X.

Remark 2.1.2 Unless stated otherwise, *throughout the following we suppose that* $\mathbb{P}(X)$ *is endowed with the weak topology*, so that $\mu_n \to \mu$ *weakly* if $\int u d\mu_n \to \int u d\mu$ for each u in $C(X)$. In this case we have that for any Borel space X:

i) $\mathbb{P}(X)$ is a Borel space. (See [11], p. 91.)

ii) If in addition X is compact, then so is $\mathbb{P}(X)$. (See [19], Theorem II 6.4.)

2.2 Multifunctions and selectors

Let X and A be (nonempty) Borel spaces.

A *multifunction* (also known as a *correspondence* or a *set-valued mapping*) F from X to A is a function such that $F(x)$ is a nonempty subset of A for each $x \in X$. A single-valued mapping $F : X \to A$ is of course an example of a multifunction.

Definition 2.2.1 a) A multifunction F from X to A is said to be *measurable* if

$$F^{-1}[U] := \{x \in X : F(x) \cap U \neq \emptyset\}$$

is a Borel subset of X for every open set $U \subset A$. The multifunction F is said to be *closed-valued* (resp. *compact-valued*) if $F(x)$ is a closed (resp. compact) set for all $x \in X$

b) The *graph* of the multifunction F is the subset of $X \times A$ defined as

$$GrF := \{(x, a) : x \in X, a \in F(x)\}$$

We say that F has a measurable graph if GrF is in $\mathcal{B}(X \times A)$.

c) $\mathbb{F}_F$ denotes the set of (single-valued) measurable functions $f : X \to A$ such that $(x, f(x))$ is in GrF, that is, $f(x) \in F(x)$ for all $x \in X$. A function $f \in \mathbb{F}_F$ is called a *selector* (or measurable selector or choice or decision function) for the multifunction F.

3 The stochastic game model

In this section we introduce the two-person zero-sum stochastic game we are interested in. We also introduce the optimality criteria, and the assumptions we shall impose throughout the remainder of this work.

We consider the two-person zero-sum game model

$$GM := \{X, A, B, \mathbb{K}_A, \mathbb{K}_B, Q, r\}$$

where :

i) X is the set of states of the game, which is assumed to be a Borel space.

ii) A and B are the action spaces for player 1 and player 2, respectively, and they are also assumed to be Borel spaces.

iii) $\mathbb{K}_A$ and $\mathbb{K}_B$ are nonempty Borel subsets of $X \times A$ and $X \times B$, respectively. For each $x \in X$ the nonempty x-section

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

of A represents the set of actions available to player 1 in state x. Analogously, the x-section $B(x) := \{b \in B : (x, b) \in \mathbb{K}_B\}$ denotes the set of actions available to player 2 in state x. Define

$$\mathbb{K} := \{(x, a, b) : x \in X, a \in A(x), b \in B(x)\},$$

which is a Borel subset of $X \times A \times B$ (see Lemma 1.1 in [15], for instance).

iv) Q is a stochastic kernel on X given $\mathbb{K}$, called the law of motion among states. If x is the state at some stage of the game and the players select actions $a \in A(x)$ and $b \in B(x)$, then $Q(\cdot|x, a, b)$ is the probability distribution of the next state of the game.

v) $r : \mathbb{K} \to \mathbb{R}$ is a measurable function 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. At each stage (or time) $t = 0, 1, \ldots$, the players 1 and 2 observe the current state $x \in X$ of the system, and then independently choose actions $a \in A(x)$ and $b \in B(x)$, respectively. As a consequence of this, the following happens: (1) player 1 receives a immediate reward $r(x, a, b)$; (2) player 2 incurres a cost $r(x, a, b)$, and (3) the system moves to a new state with distribution $Q(\cdot|x, a, b)$. Thus the goal of player 1 is to maximize his/her reward, whereas that of player 2 is to minimize his/her cost.

Let us set $\mathbb{P}_A(x) := \mathbb{P}(A(x))$ and $\mathbb{P}_B(x) := \mathbb{P}(B(x))$ for each $x \in X$. Then $x \mapsto \mathbb{P}_A(x)$ and $x \mapsto \mathbb{P}_B(x)$ define multifunctions from X to $\mathbb{P}(A)$ and from X to $\mathbb{P}(B)$, which will denoted by $\mathbb{P}_A$ and $\mathbb{P}_B$, respectively.

3.1 Strategies

Let $H_0 = X$ and $H_t = \mathbb{K} \times H_{t-1}$ for $t = 1, 2, \ldots$. For each t an element

$$h_t = (x_0, a_0, b_0, \ldots, x_{t-1}, a_{t-1}, b_{t-1}, x_t)$$

of H_t represents a "history" of the game up to time t. A randomized *strategy* π for player 1 is a sequence $\pi = \{\pi_t, t = 0, 1, \ldots\}$ of stochastic kernels π_t in $\mathbb{P}(A|H_t)$ such that

$$\pi_t(A(x_t)|h_t) = 1 \quad \forall h_t \in H_t,\ t = 0, 1, \ldots$$

We denote by Π the family of all strategies for player 1.

A strategy $\pi = \{\pi_t\}$ is called *Markov* if $\pi_t \in \mathbb{P}(A|X)$ for each $t = 0, 1, \ldots$, that is, each π_t depends only on the current state x_t of the system. The set of all Markov strategies of player 1 will be denoted by Π_M. A Markov strategy $\pi = \{\pi_t\}$ is said to be a *stationary* strategy if there exists $f \in \mathbb{P}(A|X)$ such that $\pi_t = f$ for each $t = 0, 1, \ldots$. In this case the stationary strategy π will be identified with f. We denote by Π_S the set of all stationary strategies of player 1, so that $\Pi_S \equiv \mathbb{F}_{\mathbb{P}_A}$. We have, of course,

$$\Pi_S \subset \Pi_M \subset \Pi.$$

The sets Γ, Γ_M, Γ_S of all strategies, all Markov strategies and all stationary strategies, respectively, for player 2 are defined similarly.

Let $(\Omega, \mathcal{F})$ be the (canonical) measurable space that consists of the sample space $\Omega := (X \times A \times B)^\infty$ and its product σ-algebra $\mathcal{F}$. Then for each pair of strategies $(\pi, \gamma) \in \Pi \times \Gamma$ and each "initial state" $x \in X$, by a theorem of C. Ionescu-Tulcea (see [2, p. 109], [11, p. 80]), there exists a unique probability measure $P_x^{\pi\gamma}$ and a stochastic process $\{(x_t, a_t, b_t), t = 0, 1, \ldots\}$ defined on $(\Omega, \mathcal{F})$ in a canonical way, where x_t, a_t and b_t represent the state and the actions of players 1 and 2, respectively, at each stage $t = 0, 1, \ldots$. The expectation operator with respect to $P_x^{\pi\gamma}$ is denoted by $E_x^{\pi\gamma}$

3.2 Optimality criteria

Let α be a fixed number in $(0, 1)$, and define the α-discounted expected payoff function as

$$J_\alpha(x, \pi, \gamma) := E_x^{\pi\gamma}[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t, b_t)] \tag{3.2.1}$$

for each pair of strategies (π, γ) and each initial state x. The number α is called a "discount factor".

Definition 3.2.1 For $n = 1, 2, \ldots$, we define the n-stage expected payoff function as

$$J_n(x, \pi, \gamma) := E_x^{\pi\gamma}[\sum_{t=0}^{n-1} \alpha^t r(x_t, a_t, b_t)].$$

We shall study the case in which the game model is well defined in the sense that

$$\sup_{\pi \in \Pi} \sup_{\gamma \in \Gamma} |J_\alpha(x, \pi, \gamma)| < \infty \qquad \forall x \in X. \tag{3.2.2}$$

Remark 3.2.2 The condition (3.2.2) trivially holds if r is bounded, because

$$|r(x,a,b)| \le M \text{ implies } |J_\alpha(x,\pi,\gamma)| \le \frac{M}{1-\alpha} \quad \forall\ \pi,\gamma,x.$$

Another condition that ensures (3.2.2) is given in Section 3.3, below.

To introduce our first optimality criterion we need the following concepts.

Definition 3.2.3 For each $n = 1, 2, \ldots$, the functions on X defined as

$$L_n(x) := \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} J_n(x,\pi,\gamma)$$

and

$$U_n(x) := \inf_{\gamma\in\Gamma} \sup_{\pi\in\Pi} J_n(x,\pi,\gamma)$$

are called the *lower* and the *upper* value functions, respectively, of the n-stage game.

It is clear that $L_n(\cdot) \le U_n(\cdot)$ in general, and if it holds that $L_n(x) = U_n(x)$ for all $x \in X$, then the common function is called the *value*, or value function, of the n-stage game, and is denoted by $V_n(\cdot)$.

Definition 3.2.4 Consider the n-stage game.

i) A strategy $\pi^* \in \Pi$ is said to be *optimal for player* 1 if

$$U_n(x) \le J_n(x,\pi^*,\gamma) \qquad \text{for each } \gamma\in\Gamma \text{ and } x \in X.$$

ii) A strategy $\gamma^* \in \Gamma$ is said to be *optimal for player* 2 if

$$L_n(x) \ge J_n(x,\pi,\gamma^*) \qquad \text{for each } \pi\in\Pi \text{ and } x \in X.$$

If i) and ii) hold, then (π^*,γ^*) is said to be an *optimal pair of strategies*. Equivalently, (π^*,γ^*) is an optimal pair of strategies for the n-stage game if, for all $x \in X$,

$$U_n(x) = \inf_{\gamma\in\Gamma} J_n(x,\pi^*,\gamma) \quad and \quad L_n(x) = \sup_{\pi\in\Pi} J_n(x,\pi,\gamma^*). \tag{3.2.3}$$

The value functions and the optimal pairs of strategies for $J_\alpha(x,\pi,\gamma)$ are defined similarly.

3.3 Assumptions

(a) For each state $x \in X$, the (nonempty) sets $A(x)$ and $B(x)$ of admissible actions are compact.

(b) For each $(x,a,b) \in \mathbb{K}$, $r(x,\cdot,b)$ is u.s.c. on $A(x)$, and $r(x,a,\cdot)$ is l.s.c. on $B(x)$.

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

$$\int_X v(y)Q(dy|x,\cdot,b) \quad \text{and} \quad \int_X v(y)Q(dy|x,a,\cdot)$$

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

(d) There exist a constant $M > 0$ and a measurable function $w : X \to \mathbb{R}$ such that $w(x) \geq 1$ for each $x \in X$ and

$$|r(x,a,b)| \leq Mw(x) \quad \forall\ (x,a,b) \in \mathbb{K}, \tag{3.3.4}$$

and, in addition, Assumption (c) holds when v is replaced with w.

(e) There exists a constant $1 \leq \nu < \frac{1}{\alpha}$ such that

$$\int_X w(y)Q(dy|x,a,b) \leq \nu w(x) \quad \forall\ (x,a,b) \in \mathbb{K}. \tag{3.3.5}$$

Condition (e) may be replaced with:

e') There exists a (nonempty) Borel set $C \subset X$ such that for some $\beta \in (0,1)$ and $\eta > 0$, we have

$$\int_X w(y)Q(dy|x,a,b) \leq \beta w(x) + \eta I_C(x)$$

for each $(x, a, b) \in \mathbb{K}$. Here I_C is the indicator function of the set C and w is the function introduced in (d).

4 The finite-horizon stochastic game

Our main objective in this section is to prove that the finite-horizon game has a value, and that there is an optimal pair of strategies.

4.1 Preliminaries

Let n be a positive integer. The n-stage stochastic game in which the players play up to time n is said to be a *finite-horizon game.* Let π and γ be the strategies of players 1 and 2, respectively. Then the expected payoff in that game is given by $J_n(x, \pi, \gamma)$ in Definition 3.2.1.

Before introducing the main results, we give some notation and preliminary facts. First note that the multifunctions $x \mapsto \mathbb{P}_A(x)$ and $x \mapsto \mathbb{P}_B(x)$ introduced in Section 3 are measurable and compact-valued, by Assumption 3.3(a) and Remark 2.1.2(ii).

Let $x \in X$, $\mu \in \mathbb{P}_A(x)$ and $\lambda \in \mathbb{P}_B(x)$. We define

$$r(x, \mu, \lambda) := \int_{B(x)} \int_{A(x)} r(x, a, b)\mu(da)\lambda(db), \tag{4.1.6}$$

and for every Borel set $D \subset X$

$$Q(D|x, \mu, \lambda) := \int_{B(x)} \int_{A(x)} Q(D|x, a, b)\mu(da)\lambda(db). \tag{4.1.7}$$

Definition 4.1.1 For each measurable function $u : X \to \mathbb{R}$ we define its *w-norm* as

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

where w is the function introduced in Assumption 3.3(d). We denote by $\mathbb{B}_w(X)$ the Banach space of all measurable functions u on X for which $\|u\|_w$ is finite.

For each $u \in \mathbb{B}_w(X)$ and $(x, a, b) \in \mathbb{K}$, we define

$$H(u; x, a, b) := r(x, a, b) + \alpha \int_X u(y) Q(dy|x, a, b). \qquad (4.1.8)$$

Using the notation in (4.1.6) and (4.1.7), let

$$T_\alpha u(x) := \sup_{\mu \in \mathbb{P}_A(x)} \inf_{\lambda \in \mathbb{P}_B(x)} H(u; x, \mu, \lambda) \qquad \forall\ u \in \mathbb{B}_w(X). \qquad (4.1.9)$$

Our Assumptions 3.3 and Theorem A6.3 in [2] ensure that the supremum and the infimum are indeed attained, and so the sup and inf can be replaced with max and min, respectively. Thus, we have

$$T_\alpha u(x) = \max_{\mu \in \mathbb{P}_A(x)} \min_{\lambda \in \mathbb{P}_B(x)} H(u; x, \mu, \lambda) \qquad \forall\ u \in \mathbb{B}_w(X). \qquad (4.1.10)$$

The following lemma shows that in (4.1.10) we may interchange the maximum and the minimum.

Lemma 4.1.2 *Suppose that Assumptions 3.3 hold. Then for each u in $\mathbb{B}_w(X)$:*

(a) $\quad T_\alpha u(x) = \min_{\lambda \in \mathbb{P}_B(x)} \max_{\mu \in \mathbb{P}_A(x)} H(u; x, \mu, \lambda),$

(b) there exists $f_0 \in \mathbb{F}_{\mathbb{P}_A}$ and $g_0 \in \mathbb{F}_{\mathbb{P}_B}$ such that, for all $x \in X$,

$$\begin{aligned} T_\alpha u(x) &= \max_{\mu \in \mathbb{P}_A(x)} H(u; x, \mu, g_0(x)) \\ &= \min_{\lambda \in \mathbb{P}_B(x)} H(u; x, f_0(x), \lambda) \\ &= H(u; x, f_0(x), g_0(x)), \qquad (4.1.11) \end{aligned}$$

(c) $\quad$ *$T_\alpha u$ is in $\mathbb{B}_w(X)$.*

Proof: Choose an arbitrary function u in $\mathbb{B}_w(X)$.

(a) By Assumptions 3.3(c) and the second part of 3.3(d),the integral in (4.1.8) is continuous in both $a \in A(x)$ and $b \in B(x)$ (see Lemma

8.3.7(a) in [7], for instance). This fact and Assumption 3.3(b) yield that $H(u;x,\cdot,b)$ is u.s.c. on $A(x)$, and $H(u;x,a,\cdot)$ is l.s.c. on $B(x)$. Therefore, the function $H(u;x,\mu,\lambda)$ is u.s.c. in $\mu \in \mathbb{P}_A(x)$, and l.s.c. in $\lambda \in \mathbb{P}_B(x)$; see, for example, the "extended Fatou Lemma" 8.3.7(b) and the statement (12.3.37) in [7, p. 225]. Moreover, $H(u;x,\mu,\lambda)$ is concave (as it is linear) in μ and convex in λ. Thus, by Fan's minimax theorem (see Theorem 1 in [4]), we get (a).

(b) Define

$$H_1(x,\mu) := \min_{\lambda \in \mathbb{P}_B(x)} H(u;x,\mu,\lambda) \tag{4.1.12}$$

for all $x \in X$ and $\mu \in \mathbb{P}_A(x)$. As noted in the proof of part (a), $H_1(x,\cdot)$ is u.s.c. on $\mathbb{P}_A(x)$. Therefore, by Remark 2.1.2(ii) and Theorem 1 in [10], there exists $f_0 \in \mathbb{F}_{\mathbb{P}_A}$ such that

$$H_1(x,f_0(x)) = \max_{\mu \in \mathbb{P}_A(x)} H_1(x,\mu) \quad \forall x \in X.$$

Thus, we get

$$H_1(x,f_0(x)) = \max_{\mu \in \mathbb{P}_A(x)} \min_{\lambda \in \mathbb{P}_B(x)} H(u;x,\mu,\lambda). \tag{4.1.13}$$

Hence, from (4.1.10) and (4.1.13), we have

$$T_\alpha u(x) = \min_{\lambda \in \mathbb{P}_B(x)} H(u;x,f_0(x),\lambda).$$

Similarly, if we define

$$H_2(x,\lambda) := \max_{\mu \in \mathbb{P}_A(x)} H(u;x,\mu,\lambda),$$

there exists $g_0 \in \mathbb{F}_{\mathbb{P}_B}$ such that

$$T_\alpha u(x) = \max_{\mu \in \mathbb{P}_A(x)} H(u;x,\mu,g_0(x)).$$

(c) As $|u(\cdot)| \leq \|u\|_w w(\cdot)$, from (3.3.4) and (3.3.5) we get, for any (x,a,b) in $\mathbb{K}$,

$$\begin{aligned} |H(u;x,a,b)| &\leq Mw(x) + \|u\|_w \alpha \int_X w(y) Q(dy|x,a,b) \\ &\leq (M + \alpha\nu\|u\|_w) w(x). \end{aligned} \tag{4.1.14}$$

Thus, by (4.1.11) and (4.1.14), $T_\alpha u$ is indeed in $\mathbb{B}_w(X)$. □

4.2 Existence theorem in the finite-horizon case

Theorem 4.2.1 *Suppose that Assumptions 3.3 hold. Then the finite horizon stochastic game has a value and both players have optimal Markov strategies. Moreover, if V_n is the value function for the n-stage game, then $V_n \in \mathbb{B}_w(X)$ and $V_n(x) = T_\alpha V_{n-1}(x)$ for each $n \geq 2$.*

Proof: The proof proceeds by induction. For $n = 1$ the theorem follows directly from Definition 3.2.1 and Lemma 4.1.2 with $u(\cdot) \equiv 0$. Suppose the result holds for $n-1$ ($n \geq 2$). Let $\pi_{n-1} = (f_1, f_2, \ldots, f_{n-1})$ and $\gamma_{n-1} = (g_1, g_2, \ldots, g_{n-1})$ be a pair of optimal Markov strategies for players 1 and 2, respectively, in the $(n-1)$-stage stochastic game. Then

$$V_{n-1}(\cdot) = J_{n-1}(\cdot, \pi_{n-1}, \gamma_{n-1}).$$

Let $U_n(\cdot)$ and $L_n(\cdot)$ be the upper and lower value functions, respectively; see Definition 3.2.3. Choose an arbitrary $g \in \mathbb{F}_{\mathbb{P}_B}$ and let $\gamma^g := (g, \gamma_{n-1})$. We note that, by definition of U_n,

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

Hence, for each $x \in X$,

$$U_n(x) \leq \sup_{\mu \in \mathbb{P}_A(x)} [r(x, \mu, g(x)) + \alpha \int_X \sup_{\pi \in \Pi} J_{n-1}(y, \pi, \gamma_{n-1}) Q(dy|x, \mu, g(x))],$$

that is,

$$U_n(x) \leq \sup_{\mu \in \mathbb{P}_A(x)} [r(x, \mu, g(x)) + \alpha \int_X V_{n-1}(y) Q(dy|x, \mu, g(x))].$$

Let $H_{n-1}(x, \mu, \lambda) := H(V_{n-1}; x, \mu, \lambda)$. Therefore,

$$U_n(x) \leq \sup_{\mu \in \mathbb{P}_A(x)} H_{n-1}(x, \mu, g(x))$$

and, as $g \in \mathbb{F}_{\mathbb{P}_B}$ was arbitrary,

$$U_n(x) \leq \inf_{\lambda \in \mathbb{P}_B(x)} \sup_{\mu \in \mathbb{P}_A(x)} H_{n-1}(x, \mu, \lambda)$$

for each $x \in X$. Hence, by Lemma 4.1.2, we obtain

$$U_n(x) \leq T_\alpha V_{n-1}(x).$$

Similarly, we obtain

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

From the last two inequalities, we get $U_n(x) = L_n(x) = T_\alpha V_{n-1}(x)$, so that the n-stage game has a value $V_n = T_\alpha V_{n-1}$. By Lemma 4.1.2, we now conclude that $V_n \in \mathbb{B}_w(X)$, and that there exist $f_0 \in \mathbb{F}_{\mathbb{P}_A}$ and $g_0 \in \mathbb{F}_{\mathbb{P}_B}$ such that for every $f \in \mathbb{F}_{\mathbb{P}_A}$ and $g \in \mathbb{F}_{\mathbb{P}_B}$

$$\begin{aligned} V_n(x) &= H_{n-1}(x, f_0(x), g_0(x)), \\ V_n(x) &\geq H_{n-1}(x, f(x), g_0(x)), \\ V_n(x) &\leq H_{n-1}(x, f_0(x), g(x)). \end{aligned}$$

Let $\pi_n = (f_0, f_1, \ldots, f_{n-1})$ and $\gamma_n = (g_0, g_1, \ldots, g_{n-1})$. Then it follows that π_n and γ_n are optimal for the players 1 and 2, respectively, in the n-stage stochastic game. □

5 The infinite-horizon stochastic game

In this section, we consider infinite-horizon stochastic games. We prove that the α-discounted value function V_α is a fixed point of the operator T_α in (4.1.9)-(4.1.10), that is, $V_\alpha = T_\alpha V_\alpha$, and show that the sequence $\{V_n\}$ converges geometrically to V_α in the w-norm.

5.1 Preliminaries

We consider again the Markov game model GM in Section 3 and the α-discounted expected payoff $J_\alpha(x, \pi, \gamma)$ in (3.2.1). The corresponding α-discounted lower and upper value functions are

$$\begin{aligned} L_\alpha(x) &:= \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} J_\alpha(x, \pi, \gamma), \\ U_\alpha(x) &:= \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} J_\alpha(x, \pi, \gamma). \end{aligned}$$

We wish to show that $L_\alpha(\cdot) = U_\alpha(\cdot)$, so that the value $V_\alpha(\cdot)$ for the α-discounted game exists. To do this we will first show that T_α is a contraction, which means the following.

Definition 5.1.1 Let (S, d) be a metric space. A map $T : S \to S$ is called a *contraction* if there is a number $0 \leq \tau < 1$ such that

$$d(Ts_1, Ts_2) \leq \tau d(s_1, s_2)$$

for all $s_1,\ s_2 \in S$. In this case τ is called the *modulus* of T.

Proposition 5.1.2 *(***Banach's Fixed Point Theorem.***)*
A contraction map T on a complete metric space (S, d) has a unique fixed point s^* . Moreover, $d(T^n s, s^*) \leq \tau^n d(s, s^*)$ for all $s \in S,\ n = 0, 1, \ldots,$ where τ is the modulus of T, and $T^n := T(T^{n-1})$ for $n = 1, 2, \ldots,$ with $T^0 :=$ identity.

Lemma 5.1.3 *Under Assumptions 3.3, the operator T_α defined in (4.1.9) is a contraction mapping on $\mathbb{B}_w(X)$, with modulus $\tau := \nu\alpha$ (with ν as in (3.3.5)).*

Proof: To begin, note that T_α is a monotone operator, i.e., if u and $\tilde{u}$ are functions in $\mathbb{B}_w(X)$, and $u \geq \tilde{u}$, then $T_\alpha u(x) \geq T_\alpha \tilde{u}(x)$ for all $x \in X$. On other hand, by (3.3.5), for any real number $k \geq 0$

$$T_\alpha(u + kw)(x) \leq T_\alpha u(x) + \nu\alpha k w(x) \quad \forall\ x \in X. \tag{5.1.15}$$

for all $x \in X$ and $u \in \mathbb{B}_w(X)$.

Now, to verify that T_α is a contraction, choose arbitrary u and $\tilde{u}$ in $\mathbb{B}_w(X)$. As $u \leq \tilde{u} + w\|u - \tilde{u}\|_w$, the monotonicity of T_α and (5.1.15) with $k = \|u - \tilde{u}\|_w$ yield

$$T_\alpha u(x) \leq T_\alpha(\tilde{u} + kw)(x) \leq T_\alpha \tilde{u}(x) + \nu\alpha k w(x)$$

i.e.,

$$T_\alpha u(x) - T_\alpha \tilde{u}(x) \leq \nu\alpha \|u - \tilde{u}\|_w w(x).$$

If now interchange u and $\tilde{u}$ we get

$$T_\alpha u(x) - T_\alpha \tilde{u}(x) \geq \ -\nu\alpha \|u - \tilde{u}\|_w w(x),$$

so that

$$|T_\alpha u(x) - T_\alpha \tilde{u}(x)| \leq \nu\alpha \|u - \tilde{u}\|_w w(x).$$

Hence, letting $\tau := \nu\alpha$, we obtain $\|T_\alpha u - T_\alpha \tilde{u}\|_w \leq \tau \|u - \tilde{u}\|_w$, and the lemma follows because u, $\tilde{u} \in \mathbb{B}_w(X)$ were arbitrary. □

Lemma 5.1.4 *Let M, w and ν be as in Assumptions 3.3. Moreover, let $\pi \in \Pi$ and $\gamma \in \Gamma$ be arbitrary strategies for players 1 and 2, respectively, and let $x \in X$ be an arbitrary initial state. Then for each $t = 0, 1, \ldots$*

(a) $E_x^{\pi\gamma} w(x_t) \leq \nu^t w(x)$,

(b) $|E_x^{\pi\gamma} r(x_t, a_t, b_t)| \leq M\nu^t w(x)$, *and*

(c) $\lim_{t\to\infty} \alpha^t E_x^{\pi\gamma} u(x_t) = 0$ *for all* $u \in \mathbb{B}_w(X)$.

Proof : (a) This is trivially satisfied for $t = 0$. Now, if $t \geq 1$, we have

$$\begin{aligned} E_x^{\pi\gamma}[w(x_t)|h_{t-1}, a_{t-1}, b_{t-1}] &= \int_X w(y) Q(dy|x_{t-1}, a_{t-1}, b_{t-1}) \\ &\leq \nu w(x_{t-1}) \quad \text{by } (3.3.5). \end{aligned}$$

Therefore $E_x^{\pi\gamma} w(x_t) \leq \nu E_x^{\pi\gamma} w(x_{t-1})$, which iterated yields (a).

(b) Observe that Assumption 3.3(d) yields

$$|r(x_t, a_t, b_t)| \leq M w(x_t) \qquad \forall\ t = 0, 1, \ldots,$$

so that, by (a),

$$E_x^{\pi\gamma} |r(x_t, a_t, b_t)| \leq M \nu^t w(x).$$

(c) By definition of the w-norm and part (a), we get

$$E_x^{\pi\gamma} |u(x_t)| \leq \|u\|_w E_x^{\pi\gamma} w(x_t) \leq \|u\|_w \nu^t w(x),$$

and (c) follows. □

Definition 5.1.5 Let $f \in \mathbb{F}_{\mathbb{P}_A}$, $g \in \mathbb{F}_{\mathbb{P}_B}$, and let H be as in (4.1.8). Define the operator

$$R_{fg} : \mathbb{B}_w(X) \to \mathbb{B}_w(X), \qquad u \mapsto R_{fg}u,$$

by

$$R_{fg}u(x) := H(u; x, f(x), g(x)) \qquad \forall\, x \in X. \tag{5.1.16}$$

Lemma 5.1.6 *The operator R_{fg} is a contraction operator on $\mathbb{B}_w(X)$ and $J_\alpha(x, f, g)$ is its unique fixed point in $\mathbb{B}_w(X)$.*

Proof : That R_{fg} is a contraction operator on $\mathbb{B}_w(X)$ with modulus $\tau := \alpha\nu$, follows along the same lines as the proof of Lemma 5.1.3. Therefore, R_{fg} has a unique fixed point u_{fg} in $\mathbb{B}_w(X)$, i.e.,

$$u_{fg} = R_{fg}u_{fg}. \tag{5.1.17}$$

From (5.1.17) and (5.1.16) we have then that u_{fg} is the unique solution in $\mathbb{B}_w(X)$ of the equation

$$u_{fg}(x) = r(x, f(x), g(x)) + \alpha \int_X u_{fg}(y) Q(dy|x, f(x), g(x)), \qquad \forall\, x \in X. \tag{5.1.18}$$

Moreover, iteration of (5.1.17) or (5.1.18) yields

$$u_{fg}(x) = R^n_{fg}u_{fg}(x) = E^{fg}_x[\sum_{t=0}^{n-1} \alpha^t r(x_t, f(x_t), g(x_t))] + \alpha^n E^{fg}_x u_{fg}(x_n)$$

for all $x \in X$ and $n \geq 1$, where $E^{fg}_x u(x_n) = \int_X u(y) Q^n(dy|x, f, g)$, and $Q^n(\cdot|x, f, g)$ is the n-step transition kernel of the Markov process $\{x_t\}$ when using f and g. Finally, by Lemma 5.1.4(c) and letting $n \to \infty$, we see from (3.2.1) that $u_{fg}(x) = J_\alpha(x, f, g)$ for all $x \in X$. □

5.2 Existence theorem in the infinite-horizon case

To state our main result in this section, we first recall from Theorem 4.2.1 that

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

for all $n \geq 1$ and $x \in X$, with $V_0(\cdot) \equiv 0$. That is, from Definition 3.2.3,

$$\begin{aligned} V_n(x) &= \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} J_n(x, \pi, \gamma) \\ &= \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} J_n(x, \pi, \gamma) \qquad \forall\ x \in X. \end{aligned}$$

We next consider the case when $n \to \infty$. The following theorem states, among other things, that the sequence $\{V_n\}$ converges geometrically to V_α in the w-norm.

Theorem 5.2.1 *Suppose that Assumptions 3.3 hold. Let ν be the constant in (3.3.5), and define $\tau := \nu\alpha$. Then:*

(a) The α-discounted value function V_α is the unique function in the space $\mathbb{B}_w(X)$ that satisfies the equation $T_\alpha V_\alpha = V_\alpha$, and

$$\|V_n - V_\alpha\|_w \leq M\tau^n/(1-\tau) \qquad \forall\ n = 1, 2, \ldots. \tag{5.2.19}$$

(b) There exists an optimal pair of strategies.

Proof: By Lemma 5.1.3 and Banach's Fixed Point Theorem (Proposition 5.1.2), T_α has a unique fixed point V^* in $\mathbb{B}_w(X)$, i.e.,

$$T_\alpha V^* = V^*, \tag{5.2.20}$$

and

$$\|T_\alpha^n u - V^*\|_w \leq \tau^n \|u - V^*\|_w \qquad \forall\ u \in \mathbb{B}_w(X),\ n = 0, 1, \ldots \tag{5.2.21}$$

Hence, to prove part (a) we need to show that

(i) V_α is in $\mathbb{B}_w(X)$, with norm $\|V_\alpha\|_w \leq M/(1-\tau)$, and
(ii) $V_\alpha = V^*$.

In this case, using

$$V_n = T_\alpha V_{n-1} = T_\alpha^n V_0 \qquad \forall\ n = 0, 1, \ldots, \quad V_0 = 0, \tag{5.2.22}$$

(5.2.19) will follow from (5.2.22) and (5.2.21) with $u \equiv 0$.

To prove (i), let $\pi \in \Pi$ and $\gamma \in \Gamma$ be arbitrary strategies for players 1 and 2, respectively, and let $x \in X$ be an arbitrary initial state, then (i) follows from Lemma 5.1.4(b) since a direct calculation gives

$$|J_\alpha(x,\pi,\gamma)| \leq \sum_{t=0}^{\infty} \alpha^t E|r(x_t,a_t,b_t)| \leq Mw(x)/(1-\tau)$$

with $\tau := \alpha\nu$. Thus, as $\pi \in \Pi$, $\gamma \in \Gamma$ and $x \in X$ were arbitrary,

$$|V_\alpha(x)| \leq Mw(x)/(1-\tau)$$

To prove (ii), let us note that by the equality $V^* = T_\alpha V^*$ and Lemma 4.1.2, there exists $f_* \in \mathbb{F}_{\mathbb{P}_A}$ and $g_* \in \mathbb{F}_{\mathbb{P}_B}$ such that, for all $x \in X$,

$$\begin{aligned} V^*(x) &= \sup_{\mu \in \mathbb{P}_A(x)} H(V^*; x, \mu, g_*(x)) \\ &= \inf_{\lambda \in \mathbb{P}_B(x)} H(V^*; x, f_*(x), \lambda) \qquad (5.2.23) \\ &= H(V^*; x, f_*(x), g_*(x)). \end{aligned}$$

Observe that (5.2.23) can be written as

$$V^*(x) = r(x, f_*(x), g_*(x)) + \alpha \int_X V^*(y) Q(dy|x, f_*(x), g_*(x)).$$

Then it follows from Lemma 5.1.6 that $V^*(x) = J_\alpha(x, f_*, g_*)$. Therefore, we have

$$J_\alpha(x, f_*, g_*) = \sup_{\mu \in \mathbb{P}_A(x)} [r(x,\mu,g_*(x)) + \alpha \int_X J_\alpha(y, f_*, g_*) Q(dy|x,\mu,g_*(x))]$$

for all $x \in X$. Then by standard dynamic programming results (see, for instance, [3,6,7,10,11,14]), it follows that

$$J_\alpha(x, f_*, g_*) = \sup_{\pi \in \Pi} J_\alpha(x, \pi, g_*).$$

Similarly, considering the infimum in (5.2.23) we get

$$J_\alpha(x, f_*, g_*) = \inf_{\gamma \in \Gamma} J_\alpha(x, f_*, \gamma).$$

Consequently,

$$J_\alpha(x, f_*, g_*) = \sup_{\pi\in\Pi} J_\alpha(x, \pi, g_*) \geq \inf_{\gamma\in\Gamma} \sup_{\pi\in\Pi} J_\alpha(x, \pi, \gamma),$$

and, on the other hand,

$$J_\alpha(x, f_*, g_*) = \inf_{\gamma\in\Gamma} J_\alpha(x, f_*, \gamma) \leq \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} J_\alpha(x, \pi, \gamma).$$

Hence,

$$\inf_{\gamma\in\Gamma} \sup_{\pi\in\Pi} J_\alpha(x, \pi, \gamma) = J_\alpha(x, f_*, g_*) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} J_\alpha(x, \pi, \gamma).$$

This proves that the stochastic game has a value, that the value function is $V_\alpha(x) = J_\alpha(x, f_*, g_*) = V^*(x)$ for all $x \in X$, and that f_*, g_* are optimal strategies for players 1 and 2, respectively. □

Acknowledgement

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

Francisco Ramírez-Reyes
Departamento de Matemáticas
CINVESTAV-IPN
A.P. 14-740
México D.F. 07000
México
framirez@math.cinvestav.mx

References

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

[2] R.B. Ash, Real Analysis and Probability, Academic Press, New York, 1972.

[3] D. Blackwell, *Discounted dynamic programming,* Annals of Mathematical Statistic, 36 (1965), pp. 226-235.

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

[5] M. K. Ghosh and A. Bagchi, *Stochastic games average payoff criterion,* Appl. Math. Optim. 38 (1998), pp. 283-301.

[6] O. Hernández-Lerma and J.B. Lasserre, Discrete-Time Markov Control Processes: Basic Optimality Criteria, Springer-Verlag, New York, 1996.

[7] O. Hernández-Lerma and J.B. Lasserre, Further Topics on Discrete-Time Markov Control Processes, Springer-Verlag, New York, 1999.

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

[9] O. Hernández-Lerma, J. González-Hernández and R. López-Martinez, *Constrained average cost markov control processes in Borel spaces*, CINVESTAV-IPN, Depto. de Matemáticas., reporte interno ♯ 262, 1999.

[10] C. J. Himmelberg, T. Parthasarathy and F. S. Van Vleck, *Optimal plans for dynamic programming problems,* Math. Oper. Res. 1 (1976), pp. 390-394

[11] K. Hinderer, Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter, Lecture Notes Oper. Res. and Math. Syst. 33, Springer, Berlin 1970.

[12] P. R. Kumar and T. H. Shiau, *Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games,* SIAM J. Control Optim. 19 (1981), pp. 561-585.

[13] H.-U. Küenle, *Average optimal strategies in Markov games under a geometric drift condition,* Morfimos 4 (2000), pp. 15-31.

[14] A. Maitra and T. Parthasarathy, *On stochastic games,* J. Optim. Theory Appl. 5 (1970), pp. 415-447.

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

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

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

[18] A. S. Nowak, *Universally measurables strategies in sum-zero stochastic games,* Ann. Probab. 13 (1994), pp. 269-287.

[19] K. R. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, New York, 1967.

[20] U. Rieder, *On semi-continuous dynamic games,* Preprint, Abt. für Mathematik VII, Universität Ulm, 1978.

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