KYBERNETIKA — VOLUME *55* (2019), NUMBER 1, PAGES 152–165

NASH ϵ-EQUILIBRIA FOR STOCHASTIC GAMES WITH TOTAL REWARD FUNCTIONS: AN APPROACH THROUGH MARKOV DECISION PROCESSES

Francisco J. González-Padilla and Raúl Montes-de-Oca

The main objective of this paper is to find structural conditions under which a stochastic game between two players with total reward functions has an ϵ-equilibrium. To reach this goal, the results of Markov decision processes are used to find ϵ-optimal strategies for each player and then the correspondence of a better answer as well as a more general version of Kakutani's Fixed Point Theorem to obtain the ϵ-equilibrium mentioned. Moreover, two examples to illustrate the theory developed are presented.

Keywords: stochastic games, Nash equilibrium, Markov decision processes, total rewards

Classification: 91A15, 91A50, 90C40

1. INTRODUCTION

This article presents a model of a stochastic game between two players who seek to maximize their total rewards with an infinite horizon and a numerable state space. The objective of this paper is to provide structural conditions under which an *ϵ-equilibrium* between the players exists. The main tool to achieve this is *Kakutani–Fan–Glicksberg's Fixed Point Theorem* (see [1]) (for the original version of Kakutani's Fixed Point Theorem, see [10] and [16]) which allows the authors to find, as the name suggests, the fixed points of a particular correspondence, called *correspondence of a better answer* which will turn out to be ϵ-equilibria. These fixed points will be comprised of pairs of ϵ-optimal strategies obtained through the theory of Markov decision processes (MDPs, MDP in singular) (see [13]).

Total reward problems usually require a few more assumptions than those with discounted or average rewards since the last two have a degree of control over the objective functions involved. A way in which this problem is addressed is by using a boundedness condition of the series of utilities (see Assumption 2 and Lemma 3.2 below). Along with this condition, there are structural ones which are more common in previous works such as [13], like compactness of the sets of restrictions and concavity of the utility functions.

The development of stochastic games found in the literature is usually in regard to the games with discounted or average rewards functions. There are, however, those works

DOI: 10.14736/kyb-2019-1-0152

which deal with the total reward case, such as [5], which presents this type of games and a relation to the average case with, among other things, numerable states and finite action sets. Hence, to the best of the authors' knowledge, the results of the games with a total reward criterion and compact convex actions sets in Euclidean spaces are novel and there is no previous work on this topic.

In Section 2 the basic concepts of the paper are presented which include the elements of the game and the definitions used such as total reward function and, of course, the definition of ϵ-equilibrium. In Section 3, Lemma 3.2 is provided which consists of a very useful result regarding double sequences. This lemma is the main reason for the boundedness condition mentioned above. Moreover, since Kakutani–Fan–Glicksberg's Fixed Point Theorem requires to work with sequences of strategies, a product space is constructed in which all the measures involved could be unified in different expectations in order to use both Lemma 3.2 and the Dominated Convergence Theorem. In Section 4 the main result of the paper is presented which shows the general conditions sufficient for the existence of an ϵ-equilibrium in Theorem 4.1. Finally, in Sections 5 and 6 a couple of examples are given showing how an ϵ-equilibrium exists under the assumptions given in Section 2 along with some general conclusions in which some of the problems for future work are mentioned.

2. STOCHASTIC GAMES WITH TOTAL REWARD

In this section the basic definitions are given which constitute a stochastic game between two players in which they take actions independently from each other and simultaneously. The model presented here is commonly used in works such as [9] and [15] which are based on the framework of [14], as well as the definition of a strategy for both players and when such strategies are an ϵ-equilibrium for a particular game. Now, the model is provided which describes stochastic games in a framework presented for Markov decision processes in [13], p. 28.

A *stochastic game* between two players is a tuple $\{X, A, \{A(x) : x \in X\}, B, \{B(x) : x \in X\}, Q, u_1, u_2\}$, where

- X is the Borel space (i. e. a Borel subset of a complete separable metric space) of all the states that the game can take,
- A and B are Borel spaces of all possible actions available for players 1 and 2, respectively. The non-empty measurable sets $A(x) \subseteq A$ and $B(x) \subseteq B$ are the actions available to the players when the game is at state x, and also the set

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

  is a measurable subset of $X \times A \times B$,
- $Q(\cdot|x, a, b)$, $a \in A(x), b \in B(x), x \in X$, is the transition law in every turn, and
- $u_1 : X \times A \times B \to \mathbb{R}$ and $u_2 : X \times A \times B \to \mathbb{R}$ are measurable functions used for the players' utilities.

In addition to the previous model, the following specific assumptions will be made in this article:

Assumption 1. (a) X is a numerable space endowed with the discrete topology.

(b) Both A and B are measurable subsets of the Euclidean spaces $\mathfrak{A}$ and $\mathfrak{B}$, respectively. Moreover, it is assumed that both sets $A(x)$ and $B(x)$ are compact and convex for all $x \in X$.

(c) $u_1(x,a,b)$ is concave on a for all $x \in X$ and $b \in B(x)$, and $u_2(x,a,b)$ is concave on b, for all $x \in X$ and $a \in A(x)$.

(d) u_1 and u_2 are non-negative, bounded and continuous.

2.1. Strategies

Let $x_0 = x$ be a fixed initial condition and define the sets of *feasible histories* up to time $t \geq 0$ as $H_0 = X$ and $H_t = K^t \times X = K \times H_{t-1}$, $t \geq 1$. Each element $h_t \in H_t$ is a feasible history the game can take and is of the form $h_t = (x, a_0, b_0, x_1, \ldots, a_{t-1}, b_{t-1}, x_t)$ and the set of all feasible histories is $H = K^\infty$. There exist several kinds of strategies for player 1 such as:

- A *randomized strategy* is a sequence of stochastic kernels on A given H_t for $t \geq 1$ $\{k_t\}$ such that $k_t(A(x_t)|h_t) = 1$ for all $t \geq 1$.
- A *deterministic strategy* is a sequence of measurable functions $\{f_t\}$, with $f_t : H_t \to A$, such that $f_t(h_t) \in A(x_t)$ for all $t \geq 1$.

There are also two special types of deterministic strategies that will be used.

- When $f_t : X \to A$ for all $t \geq 1$, it is called a *Markov* strategy.
- When $f_t \equiv f$ for all $t \geq 1$, it is called a *stationary* strategy and the sequence $\{f_t\}$ is denoted simply as f.

The strategies for player 2 are defined in a similar way. Now, the set of all (deterministic, Markov, stationary) strategies for player 1 and 2 will be denoted as Π_1 and Π_2, respectively. Finally, it is necessary to have a convex structure to work with, so consider the set of all convex combinations of strategies for player 1 (and player 2 respectively), that is,

$$F = \left\{ \sum_{i=1}^{n} \lambda_i f_i | \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0,\ f_i \in \Pi_1,\ n \geq 1 \right\},$$

where F is called the set of all *mixed strategies* for player 1. The set G of mixed strategies for player 2 is defined in a similar way. From this point on, whenever a strategy is mentioned, it will always mean a mixed strategy.

Remark 2.1.

(a) It is important to notice that since X is numerable and is endowed with the discrete topology, it is obtained that $F = \prod_{x \in X} A(x)$ which shows that F is in fact compact (because of the compactness of $A(x)$ for all x) with the product topology (also known as Tychonoff's topology, see Section 7, p. 93 in [11]) induced by the space A. It is also noticed that the convergence in this topology coincides with the pointwise convergence (see Definition A3.1 and Theorem A3.2, pp. 376–377 in [2] or p. 92 in [11]); this will be very useful later on when the authors will work with sequences of elements in F, that is, sequences of strategies which converge pointwise to another element in F.

(b) Observe that, since $\mathfrak{A}$ and $\mathfrak{B}$ (see Assumption 1 (b)) are locally convex (the reason of this is that $\mathfrak{A}$ and $\mathfrak{B}$ are normed spaces, with the usual norms in the Euclidean spaces), by Result (5), Section 3, p. 207 in [12] it is obtained that $\prod_{x \in X} \mathfrak{A}_x$ and $\prod_{x \in X} \mathfrak{B}_x$, where $\mathfrak{A}_x \equiv \mathfrak{A}$ and $\mathfrak{B}_x \equiv \mathfrak{B}$, are also locally convex as well as $\prod_{x \in X} (\mathfrak{A} \times \mathfrak{B})_x$, where $(\mathfrak{A} \times \mathfrak{B})_x \equiv \mathfrak{A} \times \mathfrak{B}$. Now, as $A \subseteq \mathfrak{A}$ and $B \subseteq \mathfrak{B}$ it follows that $\prod_{x \in X} (\mathfrak{A} \times \mathfrak{B})_x$ contains the set $F \times G$. This fact jointly with the locally convexity of $\prod_{x \in X} (\mathfrak{A} \times \mathfrak{B})_x$ will be relevant in the proof of Theorem 4.1.

2.2. Single player

Later in this paper the authors will find it useful to treat the previous stochastic games as convenient MDPs, in order to find ϵ-optimal strategies using the framework found in [13]. Let us see how this works: both players choose their actions independently of each other, however, a player, let's say player 1, can assume the actions the other player will take. This way, once player 2's actions (or strategy) are fixed, it is possible to consider the stochastic game with a single player, that is, an MDP (for player 1) in which:

- The state space X remains the same as in Assumption 1(a).
- $A(x)$ are the actions available to player 1 when the game is in state $x \in X$.
- The transition $Q(\cdot|x, a, b)$ and the utility function $u_1(x, a, b)$ with $x \in X$ and $a \in A(x)$ are well defined since $b \in B(x)$ is fixed (and will be of the form $b = g(x)$ for some strategy g).

And, this is similar for player 2.

Remark 2.2. The main idea behind fixing the actions of one player is to work individually with each player, find their own ϵ-optimal strategies using tools from MDPs, such as [9] and [15] and then obtain a pair of strategies which will constitute an ϵ-equilibrium with an appropriate correspondence.

2.3. Total reward

During the game, the players will receive a utility according to the action that they choose at each turn, that is, the *total reward* will be obtained by player 1 when $x_0 = x \in X$ and player 1 and 2 will choose the mixed strategies f and g in F and G, respectively, so:

$$V_1(x, f, g) = E_x^{f,g}\left[\sum_{t=0}^{\infty} u_1\left(x_t, f\left(x_t\right), g\left(x_t\right)\right)\right].$$

The function $V_2(x, f, g)$ for player 2 is also defined in a similar way.

Remark 2.3. Notice that the expectation $E_x^{f,g}$ is taken with respect to the canonical product measure ϱ obtained with the initial state $x_0 = x$ and the strategies f and g according to the Ionescu Tulcea Theorem (see Proposition C.10, p. 178 in [7]).

In this paper the following assumption will also be made:

Assumption 2. $\sup_H \sum_{t=0}^{\infty} u_1(x_t, a_t, b_t)$ and $\sup_H \sum_{t=0}^{\infty} u_2(x_t, a_t, b_t)$ are finite.

Remark 2.4. It is relevant to note that the previous assumption is also true for models with discounted rewards, see [13], Chapter 6.

2.4. Playing the simultaneous game

The game plays out as follows. Given an initial condition $x_0 \in X$, both players choose an action $a_0 \in A(x_0)$ and $b_0 \in B(x_0)$ according to their strategies f and g, then each player receives an expected reward $E_{x_0}^{f,g}[u_1(x_0, a_0, b_0)]$ and $E_{x_0}^{f,g}[u_2(x_0, a_0, b_0)]$, respectively. The game then changes to a new state $x_1 \in X$ according to the transition $Q(\cdot|x_0, a_0, b_0)$ and then the process repeats. In time, both players will receive the total of their expected profits for each action taken during the game, that is, they will receive

$$\sum_{t=0}^{\infty} E_{x_0}^{f,g}[u_1(x_t, a_t, b_t)] \text{ and } \sum_{t=0}^{\infty} E_{x_0}^{f,g}[u_2(x_t, a_t, b_t)],$$

respectively.

Remark 2.5. The previous series are well-defined thanks to Assumption 2.

2.5. ϵ-equilibrium between the players

Throughout this paper a fixed $\epsilon > 0$ will be considered.

Now, a special kind of strategies will be defined induced by the concept of ϵ-optimality in MDPs. Given $x_0 = x \in X$, a pair of strategies (f, g) is called an ϵ-*equilibrium* if

$$V_1(x, f, g) \geq \sup_{f' \in F} V_1(x, f', g) - \epsilon$$

and

$$V_2(x, f, g) \geq \sup_{g' \in G} V_2(x, f, g') - \epsilon.$$

3. PRELIMINARIES

3.1. Limit results

The main result of this paper will require the use of a series when a sequence of strategies is presented.

In the following Lemma the theorem from [6] given in Remark 3.1 will be used. The proof of this theorem in [6] although detailed is direct.

Remark 3.1. [Theorem 6.8 in Habil [6]]

Let $\{s(n,m)\}$ be a bounded double sequence of complex numbers and let $a \in \mathbb{C}$ have the property that every convergent subsequence of $\{s(n,m)\}$ converges to a. Then the sequence $\{s(n,m)\}$ converges to a.

Lemma 3.2. (a) Let $\{f_n\}$ be a sequence of strategies for player 1 which converges pointwise to the strategy f and assume that the strategy g for player 2 is fixed. If Assumptions 1 and 2 hold, then

$$\lim_{n\to\infty}\sum_{t=1}^{\infty} u_1(x_t, f_n(x_t), g(x_t)) = \sum_{t=1}^{\infty} u_1(x_t, f(x_t), g(x_t)).$$

(b) Let $\{g_n\}$ be a sequence of strategies for player 2 which converges pointwise to the strategy g and assume that the strategy f for player 1 is fixed. If Assumption 1 and 2 hold, then

$$\lim_{n\to\infty}\sum_{t=1}^{\infty} u_2(x_t, f(x_t), g_n(x_t)) = \sum_{t=1}^{\infty} u_2(x_t, f(x_t), g(x_t)).$$

P r o o f. (a) Consider the double sequence with terms

$$s(m,n) = \sum_{t=1}^{m} u_1(x_t, f_n(x_t), g(x_t))$$

and let

$$s(m_k,n_k) = \sum_{t=1}^{m_k} u_1(x_t, f_{n_k}(x_t), g(x_t))$$

be any convergent subsequence. Let L be the double limit of this subsequence. The limits $\lim_{n_k\to\infty} s(m_k,n_k)$ and $\lim_{m_k\to\infty} s(m_k,n_k)$ exist for all m_k and n_k, respectively, that is,

$$\lim_{n_k\to\infty}\sum_{t=1}^{m_k} u_1(x_t, f_{n_k}(x_t), g(x_t)) = \sum_{t=1}^{m_k} u_1(x_t, f(x_t), g(x_t))$$

and

$$\lim_{m_k\to\infty}\sum_{t=1}^{m_k} u_1(x_t, f_{n_k}(x_t), g(x_t)) = \sum_{t=1}^{\infty} u_1(x_t, f_{n_k}(x_t), g(x_t)),$$

then it follows, from Assumption 2 and Theorem 14.11, p. 141 in [3], that both iterated limits exist and are equal to the double limit $L = \sum_{t=1}^{\infty} u_1(x_t, f(x_t), g(x_t))$. This means that every convergent subsequence of the bounded sequence $\{s(m,n)\}$ converges to the same limit. By Theorem 6.8 in [6] (see Remark 3.1) the sequence $\{s(m,n)\}$ also converges to the limit L, that is,

$$\lim_{n,k\to\infty} \sum_{t=1}^{k} u_1(x_t, f_n(x_t), g(x_t)) = \lim_{n\to\infty} \sum_{t=1}^{\infty} u_1(x_t, f_n(x_t), g(x_t)) = \sum_{t=1}^{\infty} u_1(x_t, f(x_t), g(x_t)).$$

The proof for part (b) is analogous. □

3.2. Bigger spaces

Later in the paper it will be useful to work with a product space induced by a set of strategies; this way there will be a single probability and expectation that can be used for all the strategies.

Let $x \in X$ be the initial state and $\{(f_n, g_n)\}$ be a numerable set of pairs of strategies, that is, a set of paired sequences of functions defined as above and each identified as f_n and g_n. Each pair induces a probability measure $P_x^{f_n,g_n}$ obtained through the Ionescu Tulcea Theorem (see Proposition C10, p. 178 in [7]). This probability measure is the one used for the expectation $E_x^{f_n,g_n}$ in a similar way as the expectation $E_x^{f,g}$ is taken with respect to the canonical product measure ϱ in Section 2.3. Now, a new space is defined, which is obtained once again using the Ionescu Tulcea Theorem (see Proposition C10, p. 178 in [7]), $\Omega_P, \mathscr{F}_P, P$ in which

- $\Omega_P = H^\infty$
- $\mathscr{F}_P = \prod_{n=1}^{\infty} \mathscr{F}_x^{f_n,g_n}$
- $P = \prod_{n=1}^{\infty} P_x^{f_n,g_n}$,

where $\mathscr{F}_x^{f_n,g_n}$ and $P_x^{f_n,g_n}$ are the σ-algebras and probability measures of the games induced by the individual pairs (f_n, g_n).

In order to better understand what the goal of this bigger space is, let us consider a special case. Let $\{(f_n, g_n)\}$ be a sequence of strategies which converges to the pair (f,g). Each pair (f_n, g_n) and (f,g) have a product space associated with them, which are the spaces in which each game played is well-defined, namely $(H_x^{f_n,g_n}, \mathscr{F}_x^{f_n,g_n}, P_x^{f_n,g_n})$ and $(H_x^{f,g}, \mathscr{F}_x^{f,g}, P_x^{f,g})$, respectively. The next step is to construct the product space and measure induced by all these spaces given the initial state x and the transition Q. This way it is possible to consider the following:

- $\Omega_P = H_x^{f,g} \times H_x^{f_1,g_1} \times H_x^{f_2,g_2} \dots$
- $\mathscr{F}_P = \mathscr{F}_x^{f,g} \times \mathscr{F}_x^{f_1,g_1} \times \mathscr{F}_x^{f_2,g_2} \dots$
- $P = P_x^{f,g} \times P_x^{f_1,g_1} \times P_x^{f_2,g_2} \dots$,

where the space corresponding to the limit (f, g) is written at the beginning of the product but it can in fact be written anywhere with the same result.

This product measure P induces the expectation E_P which has the property that for all n,

$$E_P\left[\sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g_n\left(x_t\right)\right)\right] = E_x^{f_n, g_n}\left[\sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g_n\left(x_t\right)\right)\right],$$

as well as for the limit (f, g). This follows directly from the the Ionescu Tulcea Theorem (see Proposition C.10, p. 178 in [7]) since the previous expectation is obtained by integrating over a finite number of spaces involved, namely the space $(H_x^{f_n, g_n}, \mathscr{F}_x^{f_n, g_n}, P_x^{f_n, g_n})$ for each individual n or the space associated to the limit (f, g). The idea behind this bigger space comes when it is necessary to work with sequences of strategies that converge and it also induces a sequence of expectations, that is, let $\{f_n\}$ be a sequence of strategies which converges to f, when g is a fixed strategy and x is the initial state, then the following sequence of expectations is obtained:

$$E_x^{f_n, g}\left[\sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g\left(x_t\right)\right)\right].$$

When the limit is considered as $n \to \infty$, it is convenient to introduce the limit into the expectation, but the problem is that the probability measure used to obtain $E_x^{f_n, g}$ also depends on n. In order to solve this issue, the authors work with the product measure P which "covers" all the probability measures involved: this way it is possible to use results such as the Dominated Convergence Theorem (see [2]) to obtain the result needed in Theorem 4.1.

The previous construction considers a convergent sequence $\{(f_n, g_n)\}$ to (f, g) but the same construction can be easily adapted to consider different kinds of numerable sets. For example, if there are two sequences of strategies $\{(f_n, g_n)\}$ and $\{(f_n', g_n')\}$ each with their limits (f, g) and (f', g'), then the construction can consider a finite set of strategies involved. These cases will arise in Theorem 4.1 but the process to obtain the product spaces is very similar.

4. RESULTS: EXISTENCE OF AN ϵ-EQUILIBRIUM

In this section the main result of the paper is shown and structural conditions are given under which an ϵ-equilibrium can be guaranteed to exist in the sets of mixed strategies.

The following notation will be used for correspondences, that is, mappings from one set W to another set Z in which the image of each point x in W is a subset $\Gamma(x)$ of Z (for more information about correspondences see [1]). A correspondence T from W to Z is denoted by $T : W \rightrightarrows Z$.

Theorem 4.1. Let $x_0 = x \in X$. Under the conditions of Assumptions 1 and 2, there exists an ϵ-equilibrium for the stochastic game with total rewards on the sets of mixed strategies.

Proof. The goal is to use *Kakutani–Fan–Glicksberg's Fixed Point Theorem* (see Corollary 17.55 in [1] p. 583), to guarantee the existence of an ϵ-equilibrium. Let us assume that player 2 will choose a fixed strategy g regardless of the strategy f chosen by player 1; this transforms the two player game into a single player game in which player 1 tries to maximize a total reward MDP. This MDP will have at least one ϵ-optimal strategy according to Theorem 7.2.7 and Corollary 7.2.8 pp. 291–292 of [13] when Assumptions 1 and 2 hold. Let M_g be the set of all such ϵ-optimal strategies.

Then the previous process is repeated but now fixing a strategy f for player 1 and finding the set M_f of ϵ-optimal strategies for player 2. Using this fact, for each pair of strategies (f, g) there exists a pair of nonempty sets (M_g, M_f) of ϵ-optimal strategies for both players. Let us define a correspondence $T : F \times G \rightrightarrows F \times G$ as

$$T(f,g) = \{(f',g')|f' \in M_g,\ g' \in M_f\}.$$

Theorem 7.2.7 and Corollary 7.2.8 in [13] guarantee that there exists at least one $f' \in M_g$ and $g' \in M_f$ and therefore $T(f,g) \neq \emptyset$. It is also obtained that the set $F \times G$ is a subset of a locally convex space, see Remark 2.1 (b). Now, it is necessary to show that the correspondence T indeed has a fixed point. It is easy to verify that both sets F and G are convex because of the definition of mixed strategies. To check that they are compact, it is enough to notice that $F = \Pi_{x\in X} A(x)$ and $G = \Pi_{x\in X} B(x)$, since, from Assumption 1, $A(x)$ and $B(x)$ are compact for all x in X, so are F and G in the product topology using Tychonoff's Theorem (see p. 143 in [11]).

It is now shown that the correspondence T is convex, upper hemicontinuous and has compact images. All of the following steps must be verified for player 2 as well but, since the proof is analogous, the authors shall mainly focus on player 1. For the compactness of $T(f,g)$ a sequence $\{(f_n, g_n)\} \subset T(f,g)$ is taken which converges to some pair of strategies (f', g'), so it is necessary to show that $(f', g') \in T(f,g)$. Let E_P be the expectation obtained from the product measure as constructed above using the strategies f_n, f', g_n, g, g' with $n = 1, 2, ...,$ and $x \in X$ (see Subsection 3.2).

Since $(f_n, g_n) \in T(f,g)$, it is obtained that

$$E_P\left[\sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g\left(x_t\right)\right)\right] = E_x^{f_n,g}\left[\sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g\left(x_t\right)\right)\right] \geq \sup_{\hat{f}\in F} V_1(x,\hat{f},g) - \epsilon,$$

for all n, which implies that

$$\lim_{n\to\infty} E_P\left[\sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g\left(x_t\right)\right)\right] \geq \sup_{\hat{f}\in F} V_1(x,\hat{f},g) - \epsilon.$$

On the left hand side, by the Dominated Convergence Theorem (see [2]), it is obtained that

$$\lim_{n\to\infty} E_P\left[\sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g\left(x_t\right)\right)\right] = E_P\left[\lim_{n\to\infty} \sum_{t=0}^{\infty} u_1\left(x_t, f_n\left(x_t\right), g\left(x_t\right)\right)\right]$$

and by Lemma 3.2 (a), using the limit on n and the series itself as a double limit, it follows that

$$E_P\left[\lim_{n\to\infty}\sum_{t=0}^{\infty}u_1\left(x_t,f_n\left(x_t\right),g\left(x_t\right)\right)\right]=E_P\left[\sum_{t=0}^{\infty}\lim_{n\to\infty}u_1\left(x_t,f_n\left(x_t\right),g\left(x_t\right)\right)\right].$$

Finally, by the continuity of u_1 it is obtained that

$$E_P\left[\sum_{t=0}^{\infty}\lim_{n\to\infty}u_1\left(x_t,f_n\left(x_t\right),g\left(x_t\right)\right)\right]=E_P\left[\sum_{t=0}^{\infty}u_1\left(x_t,\lim_{n\to\infty}f_n\left(x_t\right),g\left(x_t\right)\right)\right],$$

and so

$$E_P\left[\sum_{t=0}^{\infty}u_1\left(x_t,f'\left(x_t\right),g\left(x_t\right)\right)\right]\geq\sup_{\hat{f}\in F}V_1(x,\hat{f},g)-\epsilon.$$

Next, in order to show that $T(f,g)$ is convex, consider λ, $0<\lambda<1$ and two pairs, (f',g') and $(\hat{f},\hat{g})$ in $T(f,g)$; it will be shown that the pair $(\lambda f'+(1-\lambda)\hat{f},g)$ belongs to $T(f,g)$.

Then, it is obtained that

$$\begin{aligned}E_{P'}\left[\sum_{t=0}^{\infty}u_1(x_t,c(x_t),g(x_t))\right]&\geq\lambda E_{P'}\left[\sum_{t=0}^{\infty}u_1(x_t,f'(x_t),g(x_t))\right]\\&+(1-\lambda)E_{P'}\left[\sum_{t=0}^{\infty}u_1(x_t,\hat{f}(x_t),g(x_t))\right]\\&=\lambda E_x^{f',g}\left[\sum_{t=0}^{\infty}u_1(x_t,f'(x_t),g(x_t))\right]+(1-\lambda)E_x^{\hat{f},g}\left[\sum_{t=0}^{\infty}u_1(x_t,\hat{f}(x_t),g(x_t))\right]\\&\geq\sup_{f^*\in F}V_1(x,f^*,g)-\epsilon,\end{aligned}$$

where, in this part, $E_{P'}$ is the expectation obtained from the product measure as constructed above using the strategies f', $\hat{f}$, g, $\lambda f'+(1-\lambda)\hat{f}$ and $x\in X$ (see Subsection 3.2). Therefore $T(f,g)$ is convex.

Now, it is necessary to show that T is upper hemicontinuous. To do so, consider a convergent sequence, $(f_n,g_n)\to(f,g)$ and another sequence $\{(f_n',g_n')\}$ such that $(f_n',g_n')\in T(f_n,g_n)$ for all n. It will be concluded that (f_n',g_n') has a limit point in $T(f,g)$. Since $\{(f_n',g_n')\}\subset F\times G$ and $F\times G$ is compact, there exists a subsequence $\{(f_{n_k}',g_{n_k}')\}$ which converges to the pair of strategies (f',g'). Now it will be proved that $(f',g')\in T(f,g)$ as follows (here, as above, $E_{P''}$ is the expectation obtained from the product measure using the strategies $f_n,f_n',f,f',g_n,g_n',g,g'$ with $n=1,2,...$ and $x\in X$):

$$E_{P''}\left[\sum_{t=0}^{\infty} u_1\left(x_t, f'_{n_k}\left(x_t\right), g_n\left(x_t\right)\right)\right] = E_x^{f'_{n_k}, g_n}\left[\sum_{t=0}^{\infty} u_1\left(x_t, f'_{n_k}\left(x_t\right), g_n\left(x_t\right)\right)\right]$$
$$\geq \sup_{\hat{f}\in F} V_1(x, \hat{f}, g_n) - \epsilon,$$

for all n.

This implies that

$$E_{P''}\left[\sum_{t=0}^{\infty} u_1(x_t, f'_{n_k}(x_t), g_n(x_t))\right] \geq V_1(x, \hat{f}, g_n) - \epsilon \ , \quad \text{for all } n \geq 1, \hat{f} \in F.$$

Then,

$$\lim_{n_k\to\infty} E_{P''}\left[\sum_{t=0}^{\infty} u_1(x_t, f'_{n_k}(x_t), g_{n_k}(x_t))\right] \geq \lim_{n_k\to\infty} V_1(x, \hat{f}, g_{n_k}) - \epsilon \ , \quad \text{for all } \hat{f} \in F.$$

Using arguments similar to those from the previous part, it is obtained that

$$E_{P''}\left[\sum_{t=0}^{\infty} u_1(x_t, f'(x_t), g(x_t))\right] \geq V_1(x, \hat{f}, g) - \epsilon \ , \quad \text{for all } \hat{f} \in F,$$

which, in turn, implies that

$$V_1(x, f', g) \geq \sup_{\hat{f}\in F} V_1(x, \hat{f}, g) - \epsilon.$$

Now, using Theorem 17.11 in [1] p. 561, it follows that T has a closed graph since it has been showed that it has compact, and therefore closed images and it is upper hemicontinuous. Note that Theorem 17.11 also requires the range space $F \times G$ of the correspondence T to be compact and Hausdorff, this is easily verified considering that its compactness has been proved and $F \times G$ is the product of metric spaces and therefore Hausdorff. This completes the proof of Theorem 4.1. □

Remark 4.2. The previous result makes heavy use of the existence of an ϵ-optimal strategy when the other player's own strategy is kept fixed. This approach allows to consider MDPs (single player games) and apply existing results on ϵ-optimal strategies such as their characterization as presented in works like [8] and [13], in which they give additional information regarding the strategies of the players.

5. EXAMPLES

5.1. A deterministic example

Consider two players who will have some kind of resource available to them at each period of time; they will both receive a profit depending on the amount of resource consumed. Using this information, the following game can be obtained:

- $X = \{x_t\} \subset \mathbb{R}^+$ are the amounts available to the players at each time t. Assume also that $\sum_{t=0}^{\infty} x_t < \infty$.
- $A(x) = B(x) = [0, x]$ with $x \in X$.
- $u_1(x, a, b) = a^{\alpha_1}$ and $u_2(x, a, b) = b^{\alpha_2}$ with $0 < \alpha_1, \alpha_2 < 1$, so that u_1 and u_2 are non-negative, concave and continuous.
- $Q(\{x_{t+1}\}|x_t, a_t, b_t) = 1$ for all t,

where a_t and b_t are the actions taken by players 1 and 2, respectively. Now, if a strategy g for player 2 is considered to be fixed, it is possible to obtain an MDP for player 1 with the same elements. Likewise, by fixing a strategy f for player 1 it is possible to obtain an MDP for player 2 in the same way.

Let us see that it is possible to find an ϵ-equilibrium. The aim is to find a strategy f such that $V_1(x_0, f, g) \geq \sup_{f' \in F} V_1(x_0, f', g) - \epsilon$. Let δ_1 be such that $(1 - \delta_1^{\alpha_1}) < \epsilon/\Lambda$ where $\Lambda = \sum_{t=0}^{\infty} (x_t)^{\alpha_1} < \infty$ and consider the function $f(x) = \delta_1 x$, $x \in X$. This strategy holds that

$$\sup_{f' \in F} V_1(x_0, f', g) - V_1(x_0, f, g) = \sum_{t=0}^{\infty} (x_t)^{\alpha_1} - \sum_{t=0}^{\infty} f(x_t)^{\alpha_1}$$

$$= \sum_{t=0}^{\infty} x_t^{\alpha_1} - \sum_{t=0}^{\infty} (\delta_1 x_t)^{\alpha_1} = \sum_{t=0}^{\infty} x_t^{\alpha_1} - \delta_1^{\alpha_1} x_t^{\alpha_1} = (1 - \delta_1^{\alpha_1}) \sum_{t=0}^{\infty} x_t^{\alpha_1} = (1 - \delta_1^{\alpha_1})\Lambda < \epsilon.$$

Following the same process for player 2, the strategy $g(x) = \delta_2 x$, $x \in X$ with $(1 - \delta_2^{\alpha_2}) < \epsilon/\Lambda$, is obtained. When the correspondence T is used, the couple of strategies which are ϵ-equilibria are obtained, one of which is the pair

$$f(x) = g(x) = \delta \frac{x}{2},$$

$x \in X$, where δ is such that $1 - \delta^{\alpha_i} < \epsilon/\Lambda$, $i = 1, 2$.

5.2. A stochastic example

Now, consider a game similar to the previous one but, instead of going in a fixed sequence of states, there will be a random jump from one state to another.

Let $\{\xi_t\}$ be a sequence of independent random variables where $\xi_t \sim Bin(1, \frac{1}{2})$ for all t (i. e. $P[\xi_t = 0] = P[\xi_t = 1] = \frac{1}{2}$, for all t). The game will be played the same way as above, but in this case a state space Y will be considered in which $y_0 = x_0$ and the transition $y_t = x_S$ where $S = \sum_{i=1}^{t} \xi_i + t$ will be taken into account. What this means is that the game will move along the sequence $\{x_t\}$ at either one or two steps at a time, at random. This way, all the series

$$\sum_{t=0}^{\infty} E_{x_0}^{f,g}[u_1(x_t, a_t, b_t)] \text{ and } \sum_{t=0}^{\infty} E_{x_0}^{f,g}[u_2(x_t, a_t, b_t)]$$

are bounded above by $\sum_{t=0}^{\infty} x_t^{\alpha_i}$, therefore Assumption 2 holds, and even in the worst case scenario all the states will still be visited. It then follows that an ϵ-equilibrium exists, in fact it is not unique, and one of them is still obtained like in the previous subsection, that is, one of these ϵ-equilibria is

$$f(x) = g(x) = \delta \frac{x}{2},\ x \in X.$$

6. CONCLUDING REMARKS

The main result obtained in this paper is the existence of an ϵ-equilibrium between the players under general structural conditions, some of which follow from the framework of Puterman [13] such as the numerability of the state space. The characterization of such ϵ-equilibria follows simply from results found in [13] since the strategies themselves are obtained through results on MDPs, a technique employed in works such as [9] and [15]. Other tools that were extremely helpful were those regarding double sequences found in [6] and also constructing an appropriate product space which allowed to unify the probability measures induced by all the strategies involved in the game.

Notice that the framework of the paper involves games with two players, however, the extension to consider more players is not difficult and it is straightforward.

On the other hand, some limitations still remain in the results obtained here: there are plenty of examples and applications in which the state space X is not numerable. The compactness of the restrictions is a fairly common assumption but it is possible to find examples in which an ϵ-equilibrium exists and the restrictions are not compact, so there may be work to be done in that direction. And, of course, the assumption on the supremum of the series (see Assumption 2) is indeed helpful but also a restrictive one which could be relaxed to include a wider range of examples and applications.

An additional approach which is being considered is in regard of results similar to those found in [4], which uses a risk-sensitive framework and also the objective functions associated to the total rewards used in this paper. This way the existence of an ϵ-equilibrium could also be proven under certain structural conditions for risk-sensitive games.

ACKNOWLEDGEMENT

The authors thank all the referees for their helpful comments and suggestions, which were used to aid in the improvement of this paper.

(Received December 31, 2017)

REFERENCES

[1] C. D. Aliprantis and K. C. Border: Infinite Dimensional Analysis. Springer 2006.

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

[3] R. Bartle: The Elements of Real Analysis. John Wiley and Sons, Inc. 1964. DOI:10.1002/zamm.19650450519

[4] R. Cavazos-Cadena and R. Montes-de-Oca: Optimal and nearly optimal policies in Markov decision chains with nonnegative rewards and risk-sensitive expected total-reward criterion. In: Markov Processes and Controlled Markov Chains 2002 (Z. Hou, J. A. Filar and A. Chen, eds.), Kluwer Academic Publishers, pp. 189–221. DOI:10.1007/978-1-4613-0265-0_11

[5] J. Filar and K. Vrieze: Competitive Markov Decision Processes. Springer-Verlag, New York 1997.

[6] E. D. Habil: Double sequences and double series. The Islamic Univ. J., Series of Natural Studies and Engineering *14* (2006), 1–32. (This reference is available at the Islamic University Journal's site: http://journal.iugaza.edu.ps/index.php/IUGNS/article/view/1594/1525.)

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

[8] A. Hordijk: Dynamic Programming and Markov Potential Theory. Mathematical Centre Tracts *51*, Amsterdam 1974.

[9] A. Jaśkiewicz and A. S. Nowak: Stochastic games with unbounded payoffs: Applications to robust control in Economics. Dyn. Games Appl. *1* (2011), 2, 253–279.

[10] S. Kakutani: A generalization of Brouwer's fixed point theorem. Duke Math. J. *8* (1942), 457–459. DOI:10.1215/s0012-7094-41-00838-4

[11] J. L. Kelley: General Topology. Springer, New York 1955.

[12] G. Köthe: Topological Vector Spaces I. Springer-Verlag, 1969.

[13] M. Puterman: Markov Decision Processes. John Wiley and Sons, Inc. New Jersey 1994.

[14] L. S. Shapley: Stochastic games. Proc. Nat. Acad. Sci. U. S. A. *39* (1953), 1095–1100. DOI:10.1073/pnas.39.10.1095

[15] F. Thuijsman: Optimality and Equilibria in Stochastic Games. CW1 Tract-*82*, Amsterdam 1992.

[16] E. Zeidler: Nonlinear Functional Analysis and its Applications. Springer-Verlag, New York Inc. 1988.

Francisco J. González-Padilla, Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina, 09340 Ciudad de México. México.

e-mail: morgoth80@gmail.com

Raúl Montes-de-Oca, Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina, 09340 Ciudad de México. México.

e-mail: momr@xanum.uam.mx