Extraproximal Method for Markov Chains Finite Games

S. Moya and A. Poznyak

CINVESTAV-IPN, Departamento de Control Automatico, AP 07300, Mexico, D. F., Mexico, (e-mail: smoya(apoznyak)@ctrl.cinvestav.mx)

Abstract: In this paper a regularized version of the"extraproximal method" is suggested to be applied for finding a Nash equilibrium in a multi-participant finite game where the dynamics of each player is governed by a finite controllable Markov chain. The suggested iterative technique realizes the application of a two-step procedure at each iteration: at the first (or preliminary) step some "predictive approximation" of the a current approximation is calculated; at the second step (the main step of the iteration) this prediction is used to complete the current iteration. The convergence of the suggested procedure to one of Nash-equilibrium is analyzed. The conditions guaranteeing this convergence are discussed. The numerical example demonstrates a good workability of the proposed approach.

Keywords: Matrix games; Nash equilibrium; Regularization term; Extrapoximal method.

1. INTRODUCTION

The equilibrium concept in multi-participant games was introduced by Nash [4] that may be considered as the beginning of the non-cooperative games theory generalized for players more than two [1]. In general, such type of games possesses many equilibrium points. The uniqueness of an equilibrium points is guaranteed by fulfilling of the so-called, "*strict diagonal condition*" (see [8]) which practically never takes place. So, the development of any numerical (basically, iterative) procedure providing a successful search of an equilibrium point is extremely welcomed. In [1], there is proposed an iterative version of the "*extraproximal method*" for games given on compact (non finite) sets. In [2] there is suggested to apply the "*stochastic programming approach*" which permits to represent a finite game in to a game given by its continuous variable (actually, they are "mixed strategies") defined on finite simplices. This approach operates with the corresponding *counter-coalition variables*. In these variables the original problem may be presented as a polylinear programming problem which solution is a not a simple task.

This paper consider the complete information case and follows the lead of [1] using the problem description as in [2], [3]. Here the extraproximal method is applied with an additional regularizing term that permits to state its convergence to one of Nash equilibrium points. Each player has a finite number of actions and a finite number of states. Using the randomized (mixed) strategies we may formulate the given problem as a game in variables from compact sets. The complete information on the corresponding payoff and transition matrices is assumed to be available. A numerical examples concerning two and three Markov chains non-zero sum games illustrate the suggested approach.

[1] The theory of finite two-persons games was developed in [5].

2. GAME DESCRIPTION

2.1 Homogeneous Markov chains model

Define a probability space (Ω, F, P) (Ω is a set of elementary events, F is the minimal σ-algebra of the subsets from Ω, and P is a given probabilistic measure defined for any $A \in F$) and the natural sequence $n = 1, 2, ...$, as a time argument.

Consider a collection of **controlled finite homogeneous** (with transition matrices independent on n) **Markov chains** $\mathcal{C}^k$ $(k = \overline{1,N})$ (see [7]) where each of them is a finite dynamic system described by the triplet $\mathcal{C}^k := \{X^k, U^k, \Pi^k\}$. Here:

- $X^k := \{x_1^k, x_2^k, ..., x_{K_k}^k\}$ is the finite set of states of the k^{th} Markov chain;
- $U^k := \{u_1^k, u_2^k, ..., u_{N_k}^k\}$ is the finite set of actions of the k^{th} Markov chain;
- $\Pi^k := [\Pi^{k,1} \vdots \cdots \vdots \Pi^{k,N_k}]$, $\Pi^{k,j_k} = [\pi^{k,j_k}_{s_k,s'_k}]$ is the transition matrix of $\mathcal{C}^k$ $(k = \overline{1,N})$ where the element $\pi^{k,j_k}_{s_k,s'_k}$ ($s_k, s'_k = \overline{1,K_k}$; and $j_k = \overline{1,N_k}$) represents the probability of the transition from state $x^k_{s_k}$ at time n to state $x^k_{s'_k}$ at time $n+1$ under the action $u^k_{j_k} \in U^k$, that is,

$$\pi^{k,j_k}_{s_k,s'_k} = P\left\{x^k(n+1) = x^k_{s'_k} \mid x^k(n) = x^k_{s_k}, u^k(n) = u^k_{j_k}\right\} \tag{1}$$

The random variables $x^k(n)$ are defined on the probability space (Ω, F, P) and take values in X^k.

2.2 Game Description and Randomized Strategies

Consider **N-person Markov chain game** which is the tuple

$$\left(\mathcal{C}^k, V^k, k = \overline{1,N}\right)$$

where each player is modelled by a controlled finite homogeneous Markov chain $\mathcal{C}^k$ ($k=\overline{1,N}$) and $V^k = \left(V^k_{s_1,\dots,s_N; j_1,\dots,j_N}\right)$ is a multi dimensional tensor characterizing the **payoff** associated with the k^{th} player when he selects the control action $u^k_{j_k}$ being in the state $x^k_{s_k}$ $(s_k = \overline{1,K_k}$ $j_k = \overline{1,N_k})$ and his opponents i $(i = \overline{1,N}; i \neq k)$ select the control $u^i_{j_i}$ being in state $x^i_{s_i}$.

A sequence of random stochastic matrices $D^k := \{d^k(n)\}$ is said to be **a randomized control strategy for the k^{th} player** if

- it is causal (independent on the future), that is, $d^k(n) = \left(d^{k,}_{s_k,j_k}(n)\right)_{s_k=\overline{1,K_k};j_k=\overline{1,N_k}}$ is F_{n-1} - measurable where

$$F_{n\text{-}1}:=\sigma(x^k(1),u^k(1),d^k(1);....;x^k(n\text{-}1),u^k(n\text{-}1),d^k(n\text{-}1)) \quad (2)$$

is the σ-algebra generated by $(x^k(1), u^k(1), d^k(1); ...; x^k(n-1), u^k(n-1), d^k(n-1); k = \overline{1,N})$;

- the random variables $(u^k(1),...,u^k(n-1))$ represent the "realizations" of the control actions, taking values on the finite sets $U^k = \{u^k_1,...,u^k_{N_k}\}$, and satisfy the following property:

$$d^{k,}_{s_k,j_k}(n) = \Pr\{u^k(n) = u^k_{j_k} \mid x^k(n) = x^k_{s_k} \wedge F_{n-1}\} \quad (3)$$

Denote by Σ^k the class of all **randomized strategies D^k for the k^{th} player**, that is,

$$\Sigma^k = \{\{d^k(n)\}\} \quad (4)$$

By (3), for any fixed strategy $D^k = \{d^k(n)\} \in \Sigma^k$ the conditional transition probability matrix $\Pi^k(d^k(n))$ can be defined as follows

$$\Pi^k(d^k(n)) = \left[\pi^k_{s_k,\bar{s}_k}(d(n))\right]_{s_k,\bar{s}_k=\overline{1,K_k}} \quad (5)$$

with the elements

$$\pi^k_{s_k,\bar s_k}(d(n)):=P\{x^k(n+1)=x^k_{\bar s_k} \mid x^k(n)=x^k_{s_k} \wedge F_{n-1}\} \overset{a.s.}{=} \sum_{j_k=1}^{N_k} \pi^{k,j_k}_{s_k,\bar s_k} d^{k,}_{s_k,j_k}(n) \quad (6)$$

which represents the probability to move from the states $x^k_{s_k}$ to the state $x^k_{\bar s_k}$ under the applied mixed strategy $d^{k,}_{s_k,j_k}(n)$ $(j_k=\overline{1,N_k})$.

Some important subclasses of control strategies can be considered:

(i) the class Σ^k_s of all **randomized stationary** (independent on time) **strategies** D^k for the player k, that is:

$$\Sigma^k_s = \{\{d^k(n)\} : d^k(n) = d^k\}$$

(ii) the class Σ^k_+ of all **randomized and nondegenerated stationary strategies** D^k, that is,

$$\Sigma^k_+ = \{d^k(n): d^k_{s_k,j_k}(n)=d^k_{s_k,j_k}>0\ (s_k=\overline{1,K_k}; j_k=\overline{1,N_k})\} \quad (7)$$

It is clear that $\Sigma^k_+ \subseteq \Sigma^k_s \subseteq \Sigma^k$. In the case of complete information on the payoff and transition matrices, the game is played in stages $(n = 1,2,...)$ as follows. The play starts at stage 1 in the initial states $x^k(1)$ $(k=\overline{1,N})$ which (as well as the states further realized by the process) are assumed to be completely measurable. Each of the players is allowed to randomize (with the distribution $d^k(n)$) over pure action choices $(u^k(1) \in U_k)$. These choices induce immediate payoff $V^k_{s_1,...,s_N; j_1,...,j_N}$. Each player tries to maximize the corresponding one-step payoff that he can guarantee himself by playing appropriately.

Next, the play moves to new states $(x^k(2), k = \overline{1,N})$ according to the transition probabilities $\Pi^k(d^k(n))$ $(k = \overline{1,N})$. Then, based on the obtained payoff, all participants adapt their mixed strategies constructing $d^k(n+1)$ for the next selection of the control actions and $p(n)$ is the conditional probability distribution at time n. Applied the conditional transition probability matrix (5) which changes the probability vector $p^k(n)$ to $p^k(n+1)$, that is

$$p^k(n+1) = \Pi^k(d^k(n))p^k(n)$$

2.3 Stationary Strategies and Ergodic Markov Chains

Within the class Σ^k_s of all stationary strategies and considering *ergodic* Markov chains (see [7]), for any fixed collection of stationary startegies $d^k(n) = d^k$ we have

$$p^k(n) \to p^k \text{ as } n \to \infty$$

Obviously, p^k is a functions of d^k. The *payoff function* $V^k(d)$ for the k^{th} player

$$V^k(d) := \sum_{s_1=1}^{K_1}\sum_{j_1=1}^{N_1}\cdots\sum_{s_N=1}^{K_N}\sum_{j_N=1}^{N_N} V^k_{s_1,...,s_N; j_1,...,j_N} \prod_{i=1}^{N} p^i_{s_i}(d^i)\, d^i_{s_i,j_i} \quad (8)$$

depends nonlinearly on the strategies of all the other players $d^{i\neq k}$ as well as on his own strategy d^k. Following to [7], the *individual aim* for each player can be formulated as follows:

$$V^k(d) := \sum_{s_1=1}^{K_1}\sum_{j_1=1}^{N_1}\cdots\sum_{s_N=1}^{K_N}\sum_{j_N=1}^{N_N} V^k_{s_1,...,s_N; j_1,...,j_N} \prod_{i=1}^{N} p^i_{s_i}(d^i)d^i_{s_i,j_i} \to \min_{d^k\in S^{N_k}}$$

$$d^i \in S^{N_i} := \left\{ d^i \mid \sum_{j_i=1}^{N_i} d^i_{s_i,j_i} = 1, d^i_{s_i,j_i} \geq 0 \right\}$$

$$(i = \overline{1,N}; s_i = \overline{1,K_i}, j_i = \overline{1,N_i}) \quad (9)$$

where for stationary strategies $d^i_{s_i,l_i}(n) = d^i_{s_i,l_i}$ and ergodic Markov chains the following identities hold

$$p^i_{r_i}(d^i) = \sum_{s_i=1}^{K_i}\left(\sum_{j_i=1}^{N_i}\pi^{i,j_i}_{s_i,r_i} d^i_{s_i,j_i}\right) p^i_{s_i}(d^i) \quad (10)$$

$$(i = \overline{1,N}; r_i,s_i = \overline{1,K_i}, j_i = \overline{1,N_i})$$

Sure, that the individual payoff function minimization provokes a "*conflict situation*" which can be resolved applying the game theory concept [4], [9].

2.4 Description in C-variables

The change of variable, suggested below, significantly simplifies the representation of the payoff functions converting implicit nonlinear function $V^k(d)$ in to a polylinear one. To do so, let us introduce the new variables $c^k_{s_k,j_k}$ according to the following formulas:

$$c^k_{s_k,j_k} = p^k_{s_k} d^k_{s_k,j_k}, \quad s_k = \overline{1,K_k},\ j_k = \overline{1,N_k} \quad (11)$$

Notice that for the ergodic case (see [7])

$$p^k_{s_k} = \sum_{j_k=1}^{N_k} c^k_{s_k,j_k} > 0,\ d^k_{s_k,j_k} = c^k_{s_k,j_k} \Big/ \sum_{j_k=1}^{N_k} c^k_{s_k,j_k} \quad (12)$$

The *admissible strategies* will be limited by the following requirements:

- each matrix c^k represents a stationary mixed strategy, and, hence, belongs to the *simplex* $S^{(KN)_k}$ defined by

$$S^{(KN)_k} := \left\{ c^k \in R^{K_k \cdot N_k} \mid c^k_{s_k, j_k} \geq 0, \sum_{s_k=1}^{K_k} \sum_{j_k=1}^{N_k} c^k_{s_k, j_k} = 1 \right\} \tag{13}$$

and therefore, the joint strategy matrix c belongs to the convex set S defined by

$$S := S^{(KN)_1} \times S^{(KN)_2} \times \cdots \times S^{(KN)_N} \tag{14}$$

- the joint strategy variable c satisfies the "ergodicity constraints" (10) and, hence, belongs to the convex, closed and bounded set $Q_{erg} \subset R^N$ defined by

$$\begin{gathered} Q_{erg} := Q^1_{erg} \times Q^2_{erg} \times \cdots \times Q^N_{erg} \\ Q^k_{erg} := Q^{k,1}_{erg} \times Q^{k,2}_{erg} \times \cdots \times Q^{k,K_k}_{erg} \\ Q^{k,r_k}_{erg} := \{ c := \left(c^{1,\cdot}, \ldots, c^{N,\cdot} \right) \in R^N \mid \\ \sum_{j_k=1}^{N_k} c^k_{r_k, j_k} = \sum_{s_k=1}^{K_k} \sum_{j_k=1}^{N_k} \pi^{k,j_k}_{s_k, r_k} c^k_{s_k, j_k} \} \\ r_k = \overline{1, K_k},\ k = \overline{1, N} \end{gathered} \tag{15}$$

2.5 Regularized Payoff Function Given in c-Variables

Let us introduce the following δ-regularized *payoff function* $V^k_\delta(c)$ for the k^{th} player given in variables c:

$$V^k_\delta(c) := \sum_{s_1=1}^{K_1} \sum_{j_1=1}^{N_1} \cdots \sum_{s_N=1}^{K_N} \sum_{j_N=1}^{N_N} V^k_{\substack{s_1,\ldots,s_N \\ j_1,\ldots,j_N}} \prod_{i=1}^{N} c^i_{s_i, j_i} + \frac{\delta}{2} \left\| c^k \right\|^2 \tag{16}$$

Notice that for $\delta = 0$ it coincides with (9). Using his stationary mixed strategy $d^k_{s_k, j_k} = c^k_{s_k, j_k} / \sum_{j_k=1}^{N_k} c^k_{s_k, j_k}$, each player wants to minimize his payoff (16) fulfilling the associated constraints (13) and (15).

2.6 Nash Equilibrium

Because of an arising conflict, when each player minimizes his payoff, any other ideas, how to act in this situation, are required. Here we will use the Nash-approach [4] widely used in Game Theory. A *Nash-equilibrium point* of the N-person finite game is given by a matrix c^* satisfying for all $k = \overline{1, N}$ the following equations:

$$\begin{aligned} V^k_\delta(c^*) = \min_{c^k \in S^{(KN)_k}} \{ & V^k \left(c^{1,*}, \ldots, c^{k-1,*}, c^k, c^{k+1,*}, \ldots, c^{N,*} \right) \\ & + \frac{\delta}{2} \left\| c^k \right\|^2 \mid \left(c^{1,*}, \ldots, c^{k-1,*}, c^k, c^{k+1,*}, \ldots, c^{N,*} \right) \in Q^k_{erg} \times Q^k \} \end{aligned} \tag{17}$$

Notice that the "*minimizing point*" $c^{k,*} \in S^{(KN)_k}$ is unique when another strategies $c^{j \neq k,*} \in S^{(KN)_j}$ are supposed to be fixed. The collection of equalities (17) means that, being in an equilibrium point c^*, there is no any sense for an individual participant to change unilaterally his strategy to another admissible one, since, in this case, he only increases his losses, that is, for any

$$c^k \in S^{(KN)_k} \times Q^k_{erg} \quad \left(k = \overline{1, N} \right)$$

the relation (17) is equivalent to the following systems of inequalities

$$V^k_\delta(c^*) \leq V^k_\delta \left(c^{1,*}, \ldots, c^{k-1,*}, c^k, c^{k+1,*}, \ldots, c^{N,*} \right) \quad \left(k = \overline{1, N} \right) \tag{18}$$

In spite of the fact that the individual functions $V^k_\delta(c^*)$ are strictly convex in c^k, nevertheless this does not guarantee the uniqueness of a Nash-equilibrium point. To guarantee this, the special *strict diagonal concavity condition* [8], [6] should be fulfilled which is out of the scope of this correspondence.

2.7 Joint Loss Function

Following the approach in [8] and [9], let us introduce the δ-regularized *joint Loss function* $\rho_\delta(c^*, c)$ defined by

$$\begin{gathered} \rho_\delta(c, c^*) := \sum_{k=1}^{N} V^k_\delta(c, c^*) \\ c^{|} := \left(c^1, \ldots, c^{k-1}, c^k, c^{k+1}, \ldots, c^N \right) \in S \times Q_{erg} \\ c^{*,|} := \left(c^{1,*}, \ldots, c^{k-1,*}, c^{k,*}, c^{k+1,*}, \ldots, c^{N,*} \right) \in S \times Q_{erg} \end{gathered} \tag{19}$$

for any $c \in S \times Q_{erg}$.

The fixed point $c^* \in S \times Q_{erg}$ is a Nash equilibrium (see [4]) in a N-person Markov chain game (17) if only if satisfies

$$c^* = \arg \min_{c \in S \times Q_{erg}} V^k_\delta(c, c^*) \tag{20}$$

An equilibrium point c^* in (20) can be characterized by the inequality

$$\rho_\delta(c, c^*) \geq \rho_\delta(c^*, c^*),\ \delta > 0 \tag{21}$$

Theorem 1. ([9]). A strategy $c^* \in S \times Q_{erg}$ is a Nash equilibrium point (in the sense (17)) in N-person finite Markov chain game (16) if and only if

$$\min_{c \in S \times Q_{erg}} \rho_\delta(c, c^*) = \rho_\delta(c^*, c^*),\ \delta > 0 \tag{22}$$

3. EXTRAPROXIMAL PROCEDURE FOR SOLVING EQUILIBRIUM POINT

An "*extraproximal method*", applied to resolving convex static two-person games, is designed and analyzed in [1]. Here we will apply it to solving a Nash equilibrium $c^* \in S \times Q_{erg}$, defined by (20), in Markov chain finite games (16)-(15). The idea of this method, as it may be applied to this problem, consists of the following "iterative rules" implementation:

1) The first half-step (prediction):

$$\begin{gathered} \hat{c}^k_n = \arg \min_{c^k \in S \times Q_{erg}} \left\{ \frac{1}{2} \left\| c^k - c^k_n \right\|^2 + \alpha V^k_\delta \left(c^k, c^{\hat{k}}_n \right) \right\} \\ k = \overline{1, N};\ n = 1, \ldots \end{gathered} \tag{23}$$

2) The second (basic) half-step:

$$\begin{gathered} c^k_{n+1} = \arg \min_{c^k \in S \times Q_{erg}} \left\{ \frac{1}{2} \left\| c^k - c^k_n \right\|^2 + \alpha V^k_\delta \left(c^k, \hat{c}^{\hat{k}}_n \right) \right\} \\ k = \overline{1, N};\ n = 1, \ldots \end{gathered} \tag{24}$$

where $V^k_\delta \left(c^k, \hat{c}^{\hat{k}}_n \right) = V^k_\delta \left(\hat{c}^{1,n}, \ldots, \hat{c}^{k-1,n}, c^k, \hat{c}^{k+1,n}, \ldots, \hat{c}^{N,n} \right)$ and $V^k_\delta \left(c^k, c^{\hat{k}}_n \right) = V^k(c^1_n, \ldots, c^{k-1}_n, c^k_n, c^{k+1}_n, \ldots, c^N_n)$, with the step size α from a certain fixed interval $0 < \alpha < \alpha_0$ and a small enough $\delta > 0$.

Lemma 2. Since the function $V^k_\delta \left(c^k, c^k \right)$ satisfies the condition

$$\left\| \left[V_\delta^k\left(c^k+v,c^{\hat k}+w\right) - V_\delta^k\left(c^k,c^{\hat k}+w\right)\right] - \left[V_\delta^k\left(c^k+v,c^{\hat k}\right) - V_\delta^k\left(c^k,c^{\hat k}\right)\right]\right\| \le C_k \|v\| \|w\| \tag{25}$$

for all c^k, c^k+v, $c^{\hat k}$, $c^{\hat k}+w \in S \times Q_{erg}$ and some positive constant C_k, then

$$\left\|c_{n+1}^k - \hat c_n^k\right\| \le \alpha C_k \left\|c_n^{\hat k} - \hat c_n^{\hat k}\right\| \tag{26}$$

Proof. Applied the inequality [2]

$$\frac{1}{2}\|c^* - \hat c\|^2 + \alpha V_\delta^k(c^*) \le \frac{1}{2}\|c - \hat c\|^2 + \alpha V_\delta^k(c) - \frac{1}{2}\|c - c^*\|^2$$

valid for any $c, \hat c \in S \times Q_{erg}$, we can rewrite (23)-(24) for all $c^k, c_n^k, c_{n+1}^k \in S \times Q_{erg}$ in the following manner:

$$\frac{1}{2}\left\|\hat c_n^k - c_n^k\right\|^2 + \alpha V_\delta^k\left(\hat c_n^k, c_n^{\hat k}\right) \le \frac{1}{2}\left\|c^k - c_n^k\right\|^2 + \alpha V_\delta^k\left(c^k, c_n^{\hat k}\right) - \frac{1}{2}\left\|c^k - \hat c_n^k\right\|^2 \tag{28}$$

and

$$\frac{1}{2}\left\|c_{n+1}^k - c_n^k\right\|^2 + \alpha V_\delta^k\left(c_{n+1}^k, \hat c_n^{\hat k}\right) \le \frac{1}{2}\left\|c^k - c_n^k\right\|^2 + \alpha V_\delta^k\left(c^k, \hat c_n^{\hat k}\right) - \frac{1}{2}\left\|c^k - c_{n+1}^k\right\|^2 \tag{29}$$

Let us estimate $\left\|c_{n+1}^k - \hat c_n^k\right\|$. Setting $c^k = c_{n+1}^k$ in (28) and $c^k = \hat c_n^k$ in (29), taking into account (25), and combining the resulting inequalities, we get

$$\left\|c_{n+1}^k - \hat c_n^k\right\|^2 \le \alpha \left\{\left[V_\delta^k\left(c_{n+1}^k, c_n^{\hat k}\right) - V_\delta^k\left(\hat c_n^k, c_n^{\hat k}\right)\right] - \left[V_\delta^k\left(c_{n+1}^k, \hat c_n^{\hat k}\right) - V_\delta^k\left(\hat c_n^k, \hat c_n^{\hat k}\right)\right]\right\} \le \alpha C_k \left\|c_{n+1}^k - \hat c_n^k\right\| \left\|c_n^{\hat k} - \hat c_n^{\hat k}\right\| \tag{30}$$

that leads to (26). ■

Theorem 3. Assume that there exists a solution to problem (20). Then, the sequence c_n generated by the extraproximal method (23)-(24) with the step size α, chosen from the condition $0 < \alpha < \left(\sqrt{2} C_0\right)^{-1}$ ($C_0 = \max\{C_k\}$), converges monotonically in the euclidean norm to a game equilibrium (one of possible solutions), i.e.,

$$c_n^k \to c^{k,*} \text{ as } n \to \infty$$

Proof. Setting $c^k = c^{k,*}$ in (29), then we have

$$\frac{1}{2}\left\|c_{n+1}^k - c_n^k\right\|^2 + \alpha V_\delta^k\left(c_{n+1}^k, \hat c_n^{\hat k}\right) \le \frac{1}{2}\left\|c^{k,*} - c_n^k\right\|^2 + \alpha V_\delta^k\left(c^{k,*}, \hat c_n^{\hat k}\right) - \frac{1}{2}\left\|c^{k,*} - c_{n+1}^k\right\|^2 \tag{31}$$

[2] Let us show the validity of the inequality

$$\frac{1}{2}\|c^* - \hat c\|^2 + \alpha V_\delta^k(c^*) \le \frac{1}{2}\|c - \hat c\|^2 + \alpha V_\delta^k(c) - \frac{1}{2}\|c - c^*\|^2 \quad \forall c \in S \times Q_{erg}$$

Let c^* be a minimizer of $\frac{1}{2}\|c - \hat c\|^2 + \alpha V_\delta^k(c)$ on $S \times Q_{erg}$. For this function, a necessary condition for a minimum at c^* can be written as

$$\left\langle c^* - \hat c + \alpha \nabla V_\delta^k(c^*), c - c^* \right\rangle \ge 0 \quad \forall c \in S \times Q_{erg}$$

where $\nabla V_\delta^k(c^*)$ is the gradient of $V_\delta^k(c)$. Then using the inequality

$$V_\delta^k(c) \ge V_\delta^k(c^*) + \left\langle \alpha \nabla V_\delta^k(c^*), c - c^* \right\rangle \tag{27}$$

which is true due to the convexity of $V_\delta^k(c)$, and applying the identity

$$\frac{1}{2}\|c - \hat c\|^2 = \frac{1}{2}\|c - c^*\|^2 + \langle c - c^*, c^* - \hat c\rangle + \frac{1}{2}\|c^* - \hat c\|^2$$

we obtain the desired result.

Next, setting $c^k = c_{n+1}^k$ in (28), we get

$$\frac{1}{2}\left\|\hat c_n^k - c_n^k\right\|^2 + \alpha V_\delta^k\left(\hat c_n^k, c_n^{\hat k}\right) \le \frac{1}{2}\left\|c_{n+1}^k - c_n^k\right\|^2 + \alpha V_\delta^k\left(c_{n+1}^k, c_n^{\hat k}\right) - \frac{1}{2}\left\|c_{n+1}^k - \hat c_n^k\right\|^2 \tag{32}$$

Combining the inequalities (31)-(32), we obtain

$$\left\|c^{k,*} - c_{n+1}^k\right\|^2 + \left\|c_{n+1}^k - \hat c_n^k\right\|^2 + \left\|\hat c_n^k - c_n^k\right\|^2 + 2\alpha\left[V_\delta^k\left(c_{n+1}^k, \hat c_n^{\hat k}\right) - V_\delta^k\left(\hat c_n^k, \hat c_n^{\hat k}\right) + V_\delta^k\left(\hat c_n^k, c_n^{\hat k}\right) - V_\delta^k\left(c_{n+1}^k, c_n^{\hat k}\right)\right] + 2\alpha\left\{V_\delta^k\left(\hat c_n^k, \hat c_n^{\hat k}\right) - V_\delta^k\left(c^{k,*}, \hat c_n^{\hat k}\right)\right\} \le \left\|c^{k,*} - c_n^k\right\|^2$$

The use of (30) yields

$$\left\|c^{k,*} - c_{n+1}^k\right\|^2 + \left\|c_{n+1}^k - \hat c_n^k\right\|^2 + \left\|\hat c_n^k - c_n^k\right\|^2 - 2\alpha C_k \left\|c_{n+1}^k - \hat c_n^k\right\| \left\|\hat c_n^{\hat k} - c_n^{\hat k}\right\| + 2\alpha\left\{V_\delta^k\left(\hat c_n^k, \hat c_n^{\hat k}\right) - V_\delta^k\left(c^{k,*}, \hat c_n^{\hat k}\right)\right\} \le \left\|c^{k,*} - c_n^k\right\|^2 \tag{33}$$

Using (26), we separately transform fourth part in to the inequality (33):

$$2\alpha C_k \left\|c_{n+1}^k - \hat c_n^k\right\| \left\|\hat c_n^{\hat k} - c_n^{\hat k}\right\| \le 2(\alpha C_0)^2 \left\|\hat c_n^k - c_n^k\right\|^2 \tag{34}$$

By (19), from $V_\delta^k\left(\hat c_n^k, \hat c_n^{\hat k}\right) - V_\delta^k\left(c^{k,*}, \hat c_n^{\hat k}\right)$ in (33), taking into account (21), we derive

$$V_\delta^k\left(\hat c_n^k, \hat c_n^{\hat k}\right) - V_\delta^k\left(c^{k,*}, \hat c_n^{\hat k}\right) = \rho_\delta(\hat c_n, \hat c_n) - \rho_\delta(\hat c_n, c^*) \ge 0 \tag{35}$$

Rewriting (33) applied (34)-(35), we have

$$\|c^* - c_{n+1}\|^2 + \|c_{n+1} - \hat c_n\|^2 + \left[1 - 2(\alpha C_0)^2\right]\|\hat c_n - c_n\|^2 \le \|c^* - c_n\|^2$$

Summing these inequalities from $n = 0$ up to $n = N$, we have

$$\|c_{N+1} - c^*\|^2 + \sum_{m=0}^{N}\|c_{m+1} - \hat c_m\|^2 + d\sum_{m=0}^{N}\|\hat c_m - c_m\|^2 \le \|c_0 - c^*\|^2$$

where $d = 1 - 2(\alpha C_0)^2 > 0$ by the condition of the theorem. The obtained inequality implies the bounded of the trajectories since

$$\|c_{N+1} - c^*\|^2 \le \|c_0 - c^*\|^2$$

It also implies the convergence of the series $\sum_{m=0}^{\infty}\|c_{m+1} - \hat c_m\|^2 < \infty$ and $\sum_{m=0}^{\infty}\|\hat c_m - c_m\|^2 < \infty$. Hence, $\|c_{n+1} - \hat c_n\|^2 \to 0$ and $\|\hat c_n - c_n\|^2 \to 0$ as $n \to \infty$. Since c_n is a boundedness sequence, by the Weierstrass theorem, there exists an element $\tilde c$ and a subsequence $\{c_{n_i}\}$ such that $c_{n_i} \to \tilde c$ as $n_i \to \infty$; moreover, $\|c_{n_i+1} - c_{n_i}\|^2 \to 0$ and $\|\hat c_{n_i} - c_{n_i}\|^2 \to 0$. Setting $n = n_i$ in (28), or in (29), and taking the limit on $n_i \to \infty$, we obtain

$$\tilde c^k = \arg\min_{c^k \in S \times Q_{erg}} V_\delta^k\left(c^k, \tilde c^{\hat k}\right)$$

These relations are analogous to (20), and, hence, $\tilde c^k = c^{k,*}$, i.e., any limit point of the sequence c_n^k is a solution to the problem. Since the quantity $\|c_n - c^*\|^2$ decreases monotonically, there is a unique limit point, that provides the convergence $c_n \to c^*$ as $n \to \infty$. The theorem is proven. ■

As it is follows from the structure of the cost function (16), the numerical solution, corresponding this extraproximal method, is obtained by solving the *Quadratic Programming Problem* applied to the equations (23)-(24). So, making a series of iterations, we can obtain the approximative solutions converging to one of Nash equilibriums.

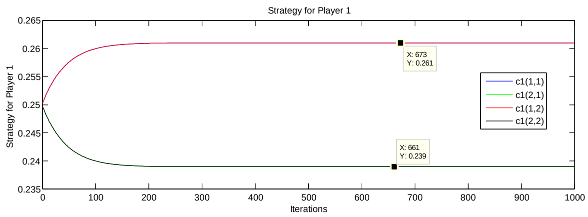

Fig. 1. Strategy for player 1

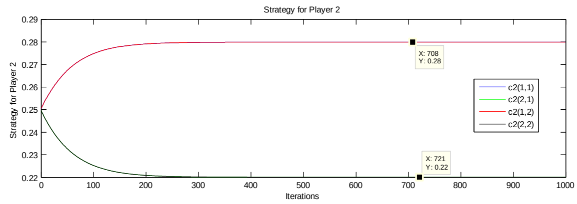

Fig. 2. Strategy for player 2

4. NUMERICAL EXAMPLE

Example 1. (Two-person game). Consider a simplest 2-person nonzero sum game with two states and two action for each player. In the matrices below the payoffs are given in the grouped format in the sequential order of players. So, the player 1 selects his actions (j_1) over the rows and the player 2 selects his actions (j_2) over the corresponding columns.

$$V^{1}_{\substack{1,1\\ j_1,j_2}} = \begin{pmatrix} -1 & 3 \\ -2 & 2 \end{pmatrix}; V^{2}_{\substack{1,1\\ j_1,j_2}} = \begin{pmatrix} -5 & 1 \\ -4 & 0 \end{pmatrix}$$

$$V^{1}_{\substack{1,2\\ j_1,j_2}} = \begin{pmatrix} 0 & -4 \\ 2 & 2 \end{pmatrix}; V^{2}_{\substack{1,2\\ j_1,j_2}} = \begin{pmatrix} 0 & 2 \\ 3 & -4 \end{pmatrix}$$

$$V^{1}_{\substack{2,1\\ j_1,j_2}} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}; V^{2}_{\substack{2,1\\ j_1,j_2}} = \begin{pmatrix} -2 & 2 \\ 3 & 0 \end{pmatrix}$$

$$V^{1}_{\substack{2,2\\ j_1,j_2}} = \begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix}; V^{2}_{\substack{2,2\\ j_1,j_2}} = \begin{pmatrix} -2 & -3 \\ -1 & -2 \end{pmatrix}$$

The transition probabilities for this game (where $j_1 = j_2$) are as follows:

$$\pi^{1,1}_{s_k,r_k} = \begin{pmatrix} 0.2 & 0.1 \\ 0.8 & 0.9 \end{pmatrix}; \pi^{2,1}_{s_k,r_k} = \begin{pmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{pmatrix}$$

$$\pi^{1,2}_{s_k,r_k} = \begin{pmatrix} 0.6 & 0.7 \\ 0.4 & 0.3 \end{pmatrix}; \pi^{2,2}_{s_k,r_k} = \begin{pmatrix} 0.7 & 0.9 \\ 0.3 & 0.1 \end{pmatrix}$$

We started with

$$c^1_0 = c^2_0 = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}$$

The solution of this Markov chain game, obtained after 300 iterations (where $\alpha = 0.001$ and $\delta = 0.01$), is as follows:

$$c^1 = \begin{pmatrix} 0.261 & 0.261 \\ 0.239 & 0.239 \end{pmatrix}, c^2 = \begin{pmatrix} 0.28 & 0.28 \\ 0.22 & 0.22 \end{pmatrix}$$

that gives

$$p^1 = p^2 = \begin{pmatrix} .5 \\ .5 \end{pmatrix}; d^1 = \begin{pmatrix} .5220 & .5220 \\ .4780 & .4780 \end{pmatrix}; d^2 = \begin{pmatrix} .5599 & .5599 \\ .4401 & .4401 \end{pmatrix}$$

The evolution of the variables through the extraproximal method are given in Fig. 1 and Fig. 2.

Example 2. (Three-person game). Here we consider the 3-person game with two states and two action for each player. In the matrices below the payoffs are given in the grouped format too in the sequential order of players. So, the player 1 selects his actions (j_1) over the rows and the player 2 selects his actions (j_2) over the corresponding columns. The actions (j_3) of the 3-rd player is supposed to be fixed.

$$V^{1}_{\substack{1,1,1\\ j_1,j_2,1}} = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix}; V^{2}_{\substack{1,1,1\\ j_1,j_2,1}} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}; V^{3}_{\substack{1,1,1\\ j_1,j_2,1}} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

$$V^{1}_{\substack{1,1,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; V^{2}_{\substack{1,1,1\\ j_1,j_2,2}} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}; V^{3}_{\substack{1,1,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

$$V^{1}_{\substack{1,1,2\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 3 \\ 0 & 4 \end{pmatrix}; V^{2}_{\substack{1,1,2\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 1 \\ 2 & 2 \end{pmatrix}; V^{3}_{\substack{1,1,2\\ j_1,j_2,1}} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$V^{1}_{\substack{1,1,2\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 2 \\ 1 & 4 \end{pmatrix}; V^{2}_{\substack{1,1,2\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 0 \\ 4 & 0 \end{pmatrix}; V^{3}_{\substack{1,1,2\\ j_1,j_2,2}} = \begin{pmatrix} 0 & -1 \\ 0 & -1 \end{pmatrix}$$

$$V^{1}_{\substack{1,2,1\\ j_1,j_2,1}} = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}; V^{2}_{\substack{1,2,1\\ j_1,j_2,1}} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}; V^{3}_{\substack{1,2,1\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 1 \\ 2 & -2 \end{pmatrix}$$

$$V^{1}_{\substack{1,2,1\\ j_1,j_2,2}} = \begin{pmatrix} 2 & 1.5 \\ 0 & 3 \end{pmatrix}; V^{2}_{\substack{1,2,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & -1 \\ 0 & 2 \end{pmatrix}; V^{3}_{\substack{1,2,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

$$V^{1}_{\substack{1,2,2\\ j_1,j_2,1}} = \begin{pmatrix} -1.5 & 0 \\ 0 & 1.5 \end{pmatrix}; V^{2}_{\substack{1,2,2\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}; V^{3}_{\substack{1,2,2\\ j_1,j_2,1}} = \begin{pmatrix} 3 & 0 \\ 4 & 0 \end{pmatrix}$$

$$V^{1}_{\substack{1,2,2\\ j_1,j_2,2}} = \begin{pmatrix} 1.5 & 2 \\ 0 & 4 \end{pmatrix}; V^{2}_{\substack{1,2,2\\ j_1,j_2,2}} = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}; V^{3}_{\substack{1,2,2\\ j_1,j_2,2}} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$V^{1}_{\substack{2,1,1\\ j_1,j_2,1}} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}; V^{2}_{\substack{2,1,1\\ j_1,j_2,1}} = \begin{pmatrix} -2 & 0 \\ -2 & 0 \end{pmatrix}; V^{3}_{\substack{2,1,1\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 2 \\ 0 & 4 \end{pmatrix}$$

$$V^{1}_{\substack{2,1,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 0 \\ -1 & 2 \end{pmatrix}; V^{2}_{\substack{2,1,1\\ j_1,j_2,2}} = \begin{pmatrix} 4 & 0 \\ 4 & 0 \end{pmatrix}; V^{3}_{\substack{2,1,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}$$

$$V^{1}_{\substack{2,1,2\\ j_1,j_2,1}} = \begin{pmatrix} 1 & 4 \\ 0 & 3 \end{pmatrix}; V^{2}_{\substack{2,1,2\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 6 \\ 1 & 4 \end{pmatrix}; V^{3}_{\substack{2,1,2\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 0 \\ 4.5 & 1 \end{pmatrix}$$

$$V^{1}_{\substack{2,1,2\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}; V^{2}_{\substack{2,1,2\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 0 \\ 3 & 4 \end{pmatrix}; V^{3}_{\substack{2,1,2\\ j_1,j_2,2}} = \begin{pmatrix} 2 & -2 \\ 4 & 2 \end{pmatrix}$$

$$V^{1}_{\substack{2,2,1\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 0 \\ 3 & 0 \end{pmatrix}; V^{2}_{\substack{2,2,1\\ j_1,j_2,1}} = \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}; V^{3}_{\substack{2,2,1\\ j_1,j_2,1}} = \begin{pmatrix} 2 & -2 \\ 4 & 2 \end{pmatrix}$$

$$V^{1}_{\substack{2,2,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 1.5 \\ 4 & 0 \end{pmatrix}; V^{2}_{\substack{2,2,1\\ j_1,j_2,2}} = \begin{pmatrix} 0 & -2 \\ 0 & -2 \end{pmatrix}; V^{3}_{\substack{2,2,1\\ j_1,j_2,2}} = \begin{pmatrix} 3 & 2 \\ 3 & 1 \end{pmatrix}$$

$$V^{1}_{\substack{2,2,2\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 0 \\ 4.5 & 0 \end{pmatrix}; V^{2}_{\substack{2,2,2\\ j_1,j_2,1}} = \begin{pmatrix} 0 & 4 \\ 2 & 4 \end{pmatrix}; V^{3}_{\substack{2,2,2\\ j_1,j_2,1}} = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$$

$$V^{1}_{\substack{2,2,2\\ j_1,j_2,2}} = \begin{pmatrix} 0 & 2 \\ 3 & 0 \end{pmatrix}; V^{2}_{\substack{2,2,2\\ j_1,j_2,2}} = \begin{pmatrix} 2 & 0 \\ 4 & 5 \end{pmatrix}; V^{3}_{\substack{2,2,2\\ j_1,j_2,2}} = \begin{pmatrix} 5 & 4 \\ 3 & 1 \end{pmatrix}$$

The transition probabilities for this game (where $j_1 = j_2 = j_3$) are as follows:

$$\pi^{1,1}_{s_k,r_k} = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}; \pi^{2,1}_{s_k,r_k} = \begin{pmatrix} 0.8 & 0.7 \\ 0.2 & 0.3 \end{pmatrix}; \pi^{3,1}_{s_k,r_k} = \begin{pmatrix} 0.9 & 0.8 \\ 0.1 & 0.2 \end{pmatrix}$$

$$\pi^{1,2}_{s_k,r_k} = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}; \pi^{2,2}_{s_k,r_k} = \begin{pmatrix} 0.2 & 0.3 \\ 0.8 & 0.7 \end{pmatrix}; \pi^{3,2}_{s_k,r_k} = \begin{pmatrix} 0.1 & 0.2 \\ 0.9 & 0.8 \end{pmatrix}$$

The solution of this Markov chain game, obtained after 600 iterations (where $\alpha = 0.001$ and $\delta = 0.01$), is as follows:

$$c^1 = \begin{pmatrix} 0.2651 & 0.2651 \\ 0.2349 & 0.2349 \end{pmatrix}, c^2 = \begin{pmatrix} 0.2325 & 0.2325 \\ 0.2675 & 0.2675 \end{pmatrix};$$

$$c^3 = \begin{pmatrix} 0.1903 & 0.1903 \\ 0.3097 & 0.3097 \end{pmatrix}$$

that gives

$$p^1 = p^2 = p^3 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, d^1 = \begin{pmatrix} 0.5302 & 0.5302 \\ 0.4698 & 0.4698 \end{pmatrix}$$

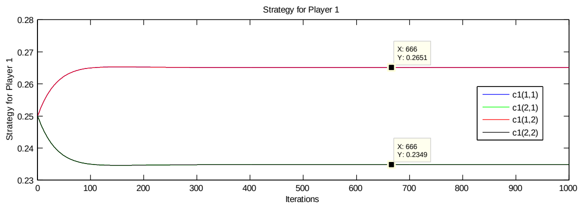

Fig. 3. Strategy for player 1

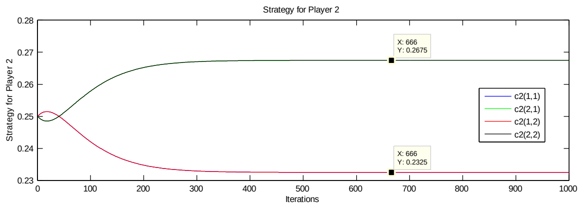

Fig. 4. Strategy for player 2

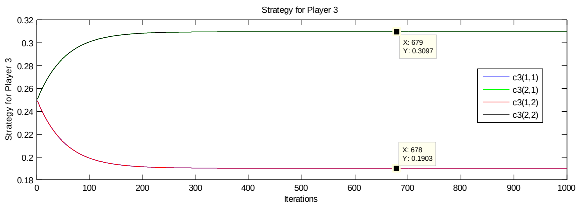

Fig. 5. Strategy for player 3

$$d^2 = \begin{bmatrix} 0.4651 & 0.4651 \\ 0.5349 & 0.5349 \end{bmatrix}, \quad d^3 = \begin{bmatrix} 0.3807 & 0.3807 \\ 0.6193 & 0.6193 \end{bmatrix}$$

The evolution of the variables through the extraproximal method are given in Fig. 3, Fig. 4 and Fig. 5.

5. CONCLUSIONS

This paper presents a new numerical procedure designed for the calculation of a Nash-equilibrium point in multi participants noncooperative Markov Chains Finites Games. The cost functions turn out to be essential nonlinear with respect to the initial control strategies d^k. We suggest to make the change of variables and to regularize the payoff function that permits to suggest the numerical procedure which demands the solution of two sequential (at each iteration) quadratic programming problem with the strictly convex (in c^k) payoffs. So, this procedure involves an iterative solution of a quadratic programming problem being considered as a natural translation of the concept of the ”*proximality*” from *Optimization Theory* to the solution of *Markov Chains Games*. The simplest numerical examples, dealing with finite two and three-person nonzero sum games, shows a good workability of the extraproximal method being applied to the solution of these problems.

REFERENCES

[1] A. S. Antipin, “An Extraproximal Method for Solving Equilibrium Programming Problems and Games”, *Computational Mathematics and Mathematical Physics* **45**, No. 11, pp. 1893-1914, 2005.

[2] M. Godoy-Alcantar, A. Poznyak and Gomez-Ramirez E., “Generalization of the Mangasarian Stone theorem for Markov Chain finite N-Person Games: LPM-approach”, *Dynamic Systems & Applications* **12**, pp. 489-508, 2003.

[3] E. Gómez-Ramírez, K. Najim and A. S. Poznyak, “Saddle-point calculation for constrained finite Markov chains”, *Journal of Economic Dynamics & Control* **27**, pp. 1833-1853, 2003.

[4] J. Nash, “Equilibrium points in n-person games”, *Proc. Nat. Acad. U.S.A*, **36**, pp. 48-49, 1950.

[5] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton Univ. Press, Princeton, NJ, 1944.

[6] A. S. Poznyak and K. Najim, “Learning through Reinforcement for N-person Repeated Game with Constraints”, *IEEE Trans. on Systems, Man, and Cybernetics*, Special issue on ”*Learning Automata Theory*”, **32**, No. 6, pp. 759-771, 2002.

[7] A. S. Poznyak, K. Najim and E. Gomez-Ramirez, Self-learning Control of Finite Markov Chains, Marcel Dekker, New York, 2000.

[8] J. B. Rosen, “Existence and uniqueness of equilibrium points for concave N-persons games,” *Econometrica* **33**, pp. 520-534, 1965.

[9] K. Tanaka and K. Yokoyama, “On ε - Equilibrium Point in a Noncooperative n - person Game”, *Journal of Mathematical Analysis and Applications*, **160**, pp. 413-423, 1991.