Proceedings of the 18th World Congress
The International Federation of Automatic Control
Milano (Italy) August 28 - September 2, 2011

Markov Decision Processes and Determining Nash Equilibria for Stochastic Positional Games

Dmitrii Lozovanu * **Stefan Pickl** ** **Erik Kropat** ***

* *Academy of Sciences of Moldova, Institute of Mathematics and Computer Science, 5 Academiei str., Chişinău, MD−2028, Moldova (e-mail: lozovanu@math.md).*
** *Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany (e-mail: stefan.pickl@unibw.de)*
*** *Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany (e-mail: erik.kropat@unibw.de)*

Abstract: A class of stochastic positional games which extend the cyclic games and Markov decision problems with average and discounted optimization costs criteria is formulated and studied. Nash equilibria conditions for considered class of stochastic positional games are derived and some approaches for determining Nash equilibria are described.

Keywords: Markov decision processes; Noncooperative Games, Stochastic positional games, Nash equilibria; Optimal stationary strategies, Cyclic games

1. INTRODUCTION

In this paper we formulate and study a class of stochastic positional games applying the game-theoretical concept to Markov decision problems with average and discounted costs optimization criteria. We consider Markov decision processes that may be controlled by several actors (players). The set of states of the system in such processes is divided into several disjoint subsets which represent the corresponding positions sets of the players. Each player has to determine which action should be taken in each state of his positions set in order to minimize his own average cost per transition or discounted expected total cost. The cost of system's transition from one state to another in the Markov process is given for each player separately. In addition the set of actions, the transition probability functions and the starting state are known. We assume that players use only stationary strategies, i.e. each player in an arbitrary his position uses the same action for an arbitrary discrete moment of time. In the considered stochastic positional games we are seeking for a Nash equilibrium.

The main results we describe in this paper are concerned with existence of Nash equilibria in the considered games and elaboration of algorithms for determining the optimal stationary strategies of players. We show that Nash equilibria for the game model with average cost payoff functions of the players exists if an arbitrary situation generated by the strategies of players induces a Markov unichain. For the game model with discounted payoff function we show that Nash equilibria always exists. The obtained results can be easy extended for antagonistic game models of Markov decision problems and the corresponding conditions for existence of saddle points in such games can be derived.

The proposed approach for Markov decision processes can be extended for multi-objective decision problems with Stackelberg and Pareto optimization principles and the corresponding algorithms for determining the optimal solutions of problems in the sense of Stackelberg and Pareto can be developed.

2. STOCHASTIC POSITIONAL GAMES WITH AVERAGE PAYOFF FUNCTIONS OF PLAYERS

We consider a class of stochastic positional games that extends and generalizes cyclic games (Gurvich [1988], Lozovanu [2006]) and Markov decision problems with average and discounted optimization costs criteria (Puterman [2005], White [1993]). The considered class of games we formulate using the framework of Markov decision process (X, A, p, c) with a finite set of states X, a finite set of actions A, a transition probability function $p : X \times X \times A \rightarrow [0, 1]$ that satisfies the condition

$$\sum_{y \in X} p^a_{x,y} = 1, \quad \forall x \in X, \quad \forall a \in A$$

and a transition cost function $c : X \times X \rightarrow \mathbb{R}$ which gives the costs $c_{x,y}$ of states transitions for the dynamical system when it makes a transition from the state $x \in X$ to another state $y \in X$.

We consider the noncooperative game model with m players in which m transition cost functions are given

$$c^i : X \times X \rightarrow \mathbb{R}, \quad i = 1, 2, \ldots, m,$$

where $c^i_{x,y}$ expresses the cost of system's transition from the state $x \in X$ to the state $y \in X$ for the player $i \in \{1, 2, \ldots, m\}$. In addition we assume that the set of states X is divided into m disjoint subsets $X_1, X_2, \ldots, X_m$

$$X = X_1 \cup X_2 \cup \cdots \cup X_m \quad (X_i \cap X_j = \emptyset, \quad \forall i \neq j),$$

where X_i represents the positions set of player $i \in \{1, 2, \ldots, m\}$. So, the Markov process is controlled by m players, where each player $i \in \{1, 2, \ldots, m\}$ fixes actions in his positions $x \in X_i$. We consider the stationary game model, i.e. we assume that each player fixes actions in the states from his positions set using stationary strategies. The stationary strategies of players we define as m maps:

$$s^1 : x \to a \in A^1(x) \quad \text{for} \quad x \in X_1;$$
$$s^2 : x \to a \in A^2(x) \quad \text{for} \quad x \in X_2;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$s^m : x \to a \in A^m(x) \quad \text{for} \quad x \in X_m,$$

where $A^i(x)$ is the set of actions of player i in the state $x \in X_i$. Without loss of generality we may consider $|A^i(x)| = |A^i| = |A|$, $\forall x \in X_i$, $i = 1, 2, \ldots, m$. In order to simplify the notation we denote the set of possible actions in a state $x \in X$ for an arbitrary player by $A(x)$.

A stationary strategy s^i, $i \in \{1, 2, \ldots, m\}$ in the state $x \in X_i$ means that at every discrete moment of time $t = 0, 1, 2, \ldots$ the player i uses the action $a = s^i(x)$. Players fix their strategy independently and do not inform each other which strategies they use in the decision process.

If the players $1, 2, \ldots, m$ fix their stationary strategies $s^1, s^2, \ldots, s^m$, respectively, then we obtain a situation $s = (s^1, s^2, \ldots, s^m)$. This situation corresponds to a simple Markov process determined by the probability distributions $p^{s^i}_{x,y}(x)$ in the states $x \in X_i$ for $i = 1, 2, \ldots, m$. We denote $P^s = (p^s_{x,y})$ the matrix of probability transitions of this Markov process. If the starting state x_{i_0} is given, then for the Markov process with the matrix of probability transitions P^s we can determine the average cost per transition $M^i_{x_0}(s^1, s^2, \ldots, s^m)$ with respect to each player $i \in \{1, 2, \ldots, m\}$ taking into account the corresponding matrix of transition costs $C^i = (c^i_{x,y})$. So, on the set of situations we can define the payoff functions of players as follows:

$$F^i_{x_{i_0}}(s^1, s^2, \ldots, s^m) = M^i_{x_{i_0}}(s^1, s^2, \ldots, s^m),\ i = 1, 2, \ldots, m.$$

In such a way we obtain a discrete noncooperative game in normal form which is determined by finite sets of strategies $\mathbb{S}^1, \mathbb{S}^2, \ldots, \mathbb{S}^m$ of m players and the payoff functions defined above. In this game we are seeking for a Nash equilibrium (Nash [2050]), i.e. we consider the problem of determining the stationary strategies

$$s^{1^*}, s^{2^*}, \ldots, s^{i-1^*}, s^{i^*}, s^{i+1^*}, \ldots, s^{m*}$$

such that

$$F^i_{x_{i_0}}(s^{1^*}, s^{2^*}, \ldots, s^{i-1^*}, s^{i^*}, s^{i+1^*} \ldots, s^{m*}) \leq$$
$$\leq F^i_{x_{i_0}}(s^{1^*}, s^{2^*}, \ldots, s^{i-1^*}, s^{i}, s^{i+1^*} \ldots, s^{m*}),$$

$(\forall s^i \in S^i, \quad i = 1, 2, \ldots, m)$. The game defined above is determined uniquely by the set of states X, the positions sets $X_1, X_2, \ldots, X_m$, the set of actions A, the cost functions $c^i : X \times X \to \mathbb{R}$, $i = 1, 2, \ldots, m$, the probability function $p : X \times X \times A \to [0, 1]$ and the starting position x_{i_0}. Therefore we denote it $(X,\ A,\ \{X_i\}_{i=\overline{1,m}},\ \{c^i\}_{i=\overline{1,m}},\ p,\ x_{i_0})$. We call this game *stochastic positional game with average payoff functions* . In the case $p^a_{x,y} = 0 \vee 1$, $\forall x, y \in X$, $\forall a \in A$ the stochastic positional game is transformed into the cyclic game studied by Gurvich [1988], Lozovanu [2009].

3. DETERMINING NASH EQUILIBRIA FOR STOCHASTIC POSITIONAL GAMES WITH AVERAGE PAYOFF FUNCTIONS

To provide the existence of Nash equilibria for the considered stochastic positional game we shall use the following condition. We assume that an arbitrary situation $s = (s^1, s^2, \ldots, s^m)$ of the game generates a Markov unichain with the corresponding matrix of probability transitions $P^s = (p^s_{x,y})$. The Markov process with such property with respect to the situations $s = (s^1, s^2, \ldots, s^m) \in S$ of the game we call *perfect Markov decision process.* We show that in this case the problem of determining Nash equilibria for a stochastic positional game can be formulated as continuous model that represents the game variant of the following optimization problem:

Minimize

$$\psi(s, q) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} s_{x,a}\, q_x \tag{1}$$

subject to

$$\begin{cases} \displaystyle\sum_{x \in X} \sum_{a \in A(x)} p^a_{x,y} s_{x,a} q_x = q_y, \quad \forall y \in X; \\ \displaystyle\sum_{x \in X} q_x = 1; \\ \displaystyle\sum_{a \in A(x)} s_{x,a} = 1, \quad \forall x \in X; \\ s_{x,a} \geq 0, \quad \forall x \in X,\ a \in A(x), \end{cases} \tag{2}$$

where

$$\mu_{x,a} = \sum_{y \in X^+(x)} c_{x,y}\, p^a_{x,y}$$

is the immediate cost in the state $x \in X$ for a fixed action $a \in A(x)$.

It is easy to observe that the problem (1), (2) represents the continuous model for Markov decision problem with average cost criterion. Indeed, an arbitrary stationary strategy $s : X \to A$ can be identified with the set of boolean variables $s_{x,a} \in \{0, 1\}$, $x \in X$, $a \in A(x)$ that satisfy the conditions

$$\sum_{a \in A(x)} s_{x,a} = 1, \quad \forall x \in X; \quad s_{x,a} \geq 0,\ \forall x \in X, a \in A.$$

These conditions determine all feasible solutions of the system (2). The rest restrictions in (2) correspond to the system of linear equations with respect to q_x for $x \in X$. This system of linear equations reflects the ergodicity condition for the limiting probability $q_x, x \in X$ in the Markov unichain, where $q_x, x \in X$ are determined uniquely for given $s_{x,a}$, $\forall x \in X, a \in A(x)$. Thus, the value of the objective function (1) expresses the average cost per transition in this Markov unichain and an arbitrary optimal solution $s^*_{x,a}, q^*_x$ $(x \in X,\ a \in A)$ of problem (1), (2) with $s^*_{x,a} \in \{0, 1\}$ represents an optimal stationary strategy for Markov decision problem with average cost criterion. If such an optimal solution is known, then an optimal action for Markov decision problem can be found by fixing $a^* = s^*(x)$ for $x \in X$ if $s^*_{x,a} = 1$.

The problem (1), (2) can be transformed into a linear programming problem using the notations $\alpha_{x,a} =$