BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 2(81), 2016, Pages 71–92
ISSN 1024–7696

Stationary Nash Equilibria for Average Stochastic Games with Finite State and Action Spaces

Dmitrii Lozovanu

Abstract. We study the problem of the existence of stationary Nash equilibria in infinite n-person stochastic games with limiting average payoff criteria for the players. The state and action spaces in the games are assumed to be finite. We present some results for the existence of stationary Nash equilibria in a multichain average stochastic game with n players. Based on constructive proof of these results we propose an approach for determining the optimal stationary strategies of the players in the case when stationary Nash equilibria in the game exist.

Mathematics subject classification: 91A15, 93E20.
Keywords and phrases: Markov decision processes, Stochastic games, Average payoffs, Stationary Nash equilibria, Optimal strategies.

1 Introduction

In this paper we investigate n-person average stochastic games with finite state and action spaces. The problem we are interested in is the existence of Nash equilibria in stationary strategies. This problem has been studied by many authors (see [4–6, 8, 9, 12, 13, 19–21]) however the existence of stationary Nash equilibria or ε-Nash equilibrium have been proved only for some classes of average stochastic games. Rogers [16] and Sobel [19] showed that stationary Nash equilibria exist for nonzero-sum stochastic games with average payoffs when the transition probability matrices induced by any stationary strategies of the players are unichain. Mertens and Neyman [12] proved the existence of uniform ε-optimal strategies in two-player zero-sum games, i.e. they showed that for every $\varepsilon > 0$ each of the two players has a strategy that guarantees the discounted value up to ε for every discount factor sufficiently close to 0. Important results for two-person non-zero sum games with average payoffs have been obtained by Vieille [20] where he shows the existence of ε-Nash equilibria. Flesch et al.[7] constructed a three-player average stochastic game with given starting state for which stationary Nash equilibria does not exist, however a cyclic Markov equilibrium for such a game exists. In general case the existence of Nash equilibria for an arbitrary stochastic game with average payoffs is an open problem. Here we formulate a condition for the existence of stationary Nash equilibria in n-person average stochastic games and based on constructive proof of this condition we propose a continuous model for the considered games that allows determining stationary Nash equilibria if such equilibria exist.

2 Formulation of average stochastic game

We present the general formulation of n-person average stochastic game and specify some basic notions that we shall use in the paper.

2.1 The framework of n-person stochastic game

A stochastic game with n players consists of the following elements:

– a state space X (which we assume to be finite);

– a finite set $A^i(x)$ of actions with respect to each player $i \in \{1, 2, \dots, n\}$ for an arbitrary state $x \in X$;

– a payoff $f^i(x, a)$ with respect to each player $i \in \{1, 2, \dots, n\}$ for each state $x \in X$ and for an arbitrary action vector $a \in \prod_i A^i(x)$;

– a transition probability function $p : X \times \prod_{x \in X} \prod_{i=1}^{n} A^i(x) \times X \rightarrow [0, 1]$ that gives the probability transitions $p^a_{x,y}$ from an arbitrary $x \in X$ to an arbitrary $y \in Y$ for a fixed action vector $a \in \prod_i A^i(x)$, where $\sum_{y \in X} p^a_{x,y} = 1, \ \forall x \in X, \ a \in \prod_i A^i(x)$;

– a starting state $x_0 \in X$.

The game starts in the state x_0 and the play proceeds in *a sequence of stages*. At *stage* t players observe state x_t and simultaneously and independently choose actions $a^i_t \in A^i(x_t), \ \ i = 1, 2, \dots, n$. Then nature selects state $y = x_{t+1}$ according to probability transitions $p^{a_t}_{x_t,y}$ for the given action vector $a_t = (a^1_t, a^2_t, \dots, a^n_t)$. Such a play of the game produces a sequence of states and actions $x_0, a_0, x_1, a_1, \dots, x_t, a_t, \dots$ that defines the corresponding stream of stage payoffs $f^1_t = f^1(x_t, a_t), \ f^2_t = f^2(x_t, a_t), \dots, \ f^n_t = f^n(x_t, a_t), \ \ t = 0, 1, 2, \dots$. *The infinite average stochastic game* is the game with payoffs of players

$$\omega^i_{x_0} = \lim_{t \to \infty} \inf \ \mathsf{E}\left(\frac{1}{t} \sum_{\tau=0}^{t-1} f^i_\tau \right), \quad i = 1, 2, \dots, n,$$

where $\omega^i_{x_o}$ expresses the *average payoff per transition* of player i in infinite game. In the case $n = 1$ this game becomes the average Markov decision problem with a transition probability function $p : X \times \prod_{x \in X} A(x) \times X \rightarrow [0, 1]$ and immediate rewards $f(x, a) = f^1(x, a)$ in the states $x \in X$ for given actions $a \in A(x) = A^1(x)$.

In the paper we will study the stochastic games when players use pure and mixed stationary strategies of selection of the actions in the states.

2.2 Pure and mixed stationary strategies of the players

A strategy of player $i \in \{1, 2, \ldots, n\}$ in a stochastic game is a mapping s^i that for every state $x_t \in X$ provides a probability distribution over the set of actions $A^i(x_t)$. If these probabilities take only values 0 and 1, then s^i is called *a pure strategy*, otherwise s^i is called *a mixed strategy.* If these probabilities depend only on the state $x_t = x \in X$ (i. e. s^i do not depend on t), then s^i is called *a stationary strategy.* This means that *a pure stationary strategy* of player $i \in \{1, 2, \ldots, n\}$ can be regarded as a map

$$s^i : x \to a^i \in A^i(x) \ \text{ for } \ x \in X$$

that determines for each state x an action $a^i \in A^i(x)$, i.e. $s^i(x) = a^i$. Obviously, the corresponding sets of pure stationary strategies $S^1, S^2, \ldots, S^n$ of the players in the game with finite state and action spaces are finite sets.

In the following we will identify a pure stationary strategy $s^i(x)$ of player i with the set of boolean variables $s^i_{x,a^i} \in \{0, 1\}$, where for a given $x \in X$ $s^i_{x,a^i} = 1$ if and only if player i fixes the action $a^i \in A^i(x)$. So, we can represent the set of pure stationary strategies S^i of player i as the set of solutions of the following system:

$$\begin{cases} \sum\limits_{a^i \in A^i(x)} s^i_{x,a^i} = 1, & \forall x \in X; \\ s^i_{x,a^i} \in \{0, 1\}, & \forall x \in X, \ \ \forall a^i \in A^i(x). \end{cases}$$

If in this system we change the restriction $s^i_{x,a^i} \in \{0, 1\}$ for $x \in X$, $a^i \in A^i(x)$ by the condition $0 \le s^i_{x,a^i} \le 1$ then we obtain the set of stationary strategies in the sense of Shapley [17], where s^i_{x,a^i} is treated as the probability of choices of the action a^i by player i every time when the state x is reached by any route in the dynamic stochastic game. Thus, we can identify the set of mixed stationary strategies of the players with the set of solutions of the system

$$\begin{cases} \sum\limits_{a^i \in A^i(x)} s^i_{x,a^i} = 1, & \forall x \in X; \\ s^i_{x,a^i} \ge 0, & \forall x \in X, \ \ \forall a^i \in A^i(x) \end{cases} \tag{1}$$

and for a given profile $s = (s^1, s^2, \ldots, s^n)$ of mixed strategies $s^1, s^2, \ldots, s^n$ of the players the probability transition $p^s_{x,y}$ from a state x to a state y can be calculated as follows

$$p^s_{x,y} = \sum_{(a^1, a^2, \ldots, a^n) \in A(x)} \prod_{k=1}^{n} s^k_{x,a^k} p^{(a^1, a^2, \ldots, a^n)}_{x,y}. \tag{2}$$

In the sequel we will distinguish stochastic games in pure and mixed stationary strategies.

2.3 Average stochastic games in pure stationary strategies

Let $s = (s^1, s^2, \ldots, s^n)$ be a profile of pure stationary strategies of the players and denote by $a(s) = (a^1(s), a^2(s), \ldots, a^n(s)) \in \prod\limits_{x \in X} \prod\limits_{i=1}^{n} A^i(x)$ the action vector that corresponds to s and determines the probability distributions $p^s_{x,y} = p^{a(s)}_{x,y}$ in the states $x \in X$. Then the average payoffs per transition $\omega^1_{x_0}(s),\ \omega^2_{x_0}(s), \dots, \omega^n_{x_0}(s)$ for the players are determined as follows

$$\omega^i_{x_0}(s) = \sum_{y \in X} q^s_{x_0,y} f^i(y, a(s)), \quad i = 1, 2, \dots, n,$$

where $q^s_{x_o,y}$ represent the limiting probabilities in the states $y \in X$ for the Markov process with probability transition matrix $P^s = (p^s_{x,y})$ when the transitions start in x_0. So, if for the Markov process with probability matrix P^s the corresponding limiting probability matrix $Q^s = (q^s_{x,y})$ is known then $\omega^1_x, \omega^2_x, \dots, \omega^n_x$ can be determined for an arbitrary starting state $x \in X$ of the game. The functions $\omega^1_{x_0}(s),\ \omega^2_{x_0}(s), \dots, \omega^n_{x_0}(s)$ on $S = S^1 \times S^2 \times \cdots \times S^n$ define a game in normal form that corresponds to an infinite *average stochastic game in pure stationary strategies*. This game is determined by the set of states X, the sets of actions of the players $\{A^i(x)\}_{i=\overline{1,n}}$, the probability function p, the set of stage payoffs $\{f^i(x,a\}_{i=\overline{1,n}}$ and the starting position of the game x_0. Therefore we denote this game by $(X, \{A^i(x)\}_{i=\overline{1,n}}, \{f^i(x,a\}_{i=\overline{1,n}}, p, x_0)$. If the starting position of the game is chosen randomly according to distribution probabilities $\{\theta_x\}$ in X then such a game we denote $(X, \{A^i(x)\}_{i=\overline{1,n}}, \{f^i(x,a\}_{i=\overline{1,n}}, p, \{\theta_x\})$.

If an arbitrary profile $s = (s^1, s^2, \dots, s^n)$ of pure stationary strategies in a stochastic game induces a probability matrix P^s that corresponds to a Markov unichain then we say that the game possesses the unichain property and shortly we call it *unichain stochastic game;* otherwise we call it *multichain stochastic game.*

For an average stochastic game in pure strategies a Nash equilibrium may not exist. Therefore in this paper we study stochastic games in the case when players use mixed stationary strategies.

2.4 Stochastic games in mixed stationary strategies

Let $s = (s^1, s^2, \dots, s^n)$ be a profile of mixed stationary strategies of the players. Then elements of probability transition matrix $P^s = (p^s_{x,y})$ in the Markov process induced by s can be calculated according to (2). Therefore if $Q^s = (q^s_{x,y})$ is the limiting probability matrix of P^s then the average payoffs per transition $\omega^1_{x_0}(s),\ \omega^2_{x_0}(s), \dots, \omega^n_{x_0}(s)$ for the players are determined as follows

$$\omega^i_{x_0}(s) = \sum_{y \in X} q^s_{x_0,y} f^i(y, s), \quad i = 1, 2, \dots, n,$$

where

$$f^i(y,s) = \sum_{(a^1,a^2,\dots,a^n) \in A(y)} \prod_{k=1}^{n} s^k_{y,a^k} f^i(y, a^1, a^2, \dots, a^n)$$

expresses the average payoff (immediate reward) in the state $y \in X$ of player i when the corresponding stationary strategies $s^1, s^2, \dots, s^n$ have been applied by players $1, 2, \dots, n$ in y.

Let $\overline{S}^1, \overline{S}^2, \dots, \overline{S}^n$ be the corresponding sets of mixed stationary strategies for the players $1, 2, \dots, n$, i.e. each $\overline{S}^i$ for $i \in \{1, 2, \dots, n\}$ represents the set of solutions of system (1). Then the functions $\omega^1_{x_0}(s)$, $\omega^2_{x_0}(s), \dots, \omega^n_{x_0}(s)$ on $\overline{S} = \overline{S}^1 \times \overline{S}^2 \times \dots \times \overline{S}^n$, define a game in normal form. This game corresponds to an infinite *average stochastic game in mixed stationary strategies.*

3 Preliminaries

We present some results for the average Markov decision problem and for the average stochastic game with unichain property that we shall use for the multichain average stochastic games.

3.1 A continuous model for the average Markov decision problem with unichain property

In [9] it has been shown that an average Markov decision problem with unichain property can be formulated as the following optimization problem:
Maximize

$$\psi(s,q) = \sum_{x\in X} \sum_{a\in A(x)} f(x,a) s_{x,a}\, q_x \tag{3}$$

subject to

$$\begin{cases} q_y - \sum\limits_{x\in X} \sum\limits_{a\in A(x)} p^a_{x,y} s_{x,a} q_x = 0, & \forall y \in X; \\ \sum\limits_{x\in X} q_x = 1; & \\ \sum\limits_{a\in A(x)} s_{x,a} = 1, & \forall x \in X; \\ s_{x,a} \geq 0, \ \ \forall x \in X, \ a \in A(x). & \end{cases} \tag{4}$$

Here $f(x,a)$ represents the immediate reward in the state $x \in X$ for a given action $a \in A(x)$ in the unichain problem and $p^a_{x,y}$ expresses the probability transition from $x \in X$ to $y \in X$ for $a \in A(x)$. The variables $s_{x,a}$ correspond to strategies of selection of the actions $a \in A(x)$ in the states $x \in X$ and q_x for $x \in X$ represent the corresponding limiting probabilities in the states $x \in X$ for the probability transition matrix $P^s = (p^s_{x,y})$ induced by stationary strategy s.

In this problem the average reward $\psi(s,q)$ is maximized under the conditions (4) that determines the set of feasible stationary strategies in the unichain problem. An optimal solution (s^*, q^*) of problem (3), (4) with $s^*_{x,a} \in \{0,1\}$ corresponds to an optimal stationary strategy $s^* : X \to A$ where $a^* = s^*(x)$ for $x \in X$ if $s^*_{x,a} = 1$. Using the notations $\alpha_{x,a} = s_{x,a} q_x$, for $x \in X, a \in A(x)$, problem (3), (4) can be easily transformed into the following linear programming problem:
Maximize

$$\overline{\psi}(\alpha) = \sum_{x\in X} \sum_{a\in A(x)} f(x,a)\, \alpha_{x,a} \tag{5}$$

subject to

$$\begin{cases} q_y - \sum\limits_{x\in X}\sum\limits_{a\in A(x)} p^a_{x,y}\alpha_{x,a} = 0, & \forall y\in X;\\ \sum\limits_{x\in X} q_x = 1; & \\ \sum\limits_{a\in A(x)} \alpha_{x,a} - q_x = 0, & \forall x\in X;\\ \alpha_{x,a}\ge 0, \ \ \forall x\in X,\ a\in A(x). & \end{cases} \tag{6}$$

This problem can be simplified by eliminating q_x from (6) and finally we obtain the problem in which it is necessary to maximize the objective function (5) on the set of solutions of the following system:

$$\begin{cases} \sum\limits_{a\in A(y)} \alpha_{y,a} - \sum\limits_{x\in X}\sum\limits_{a\in A(x)} p^a_{x,y}\,\alpha_{x,a} = 0, \quad \forall y\in X;\\ \sum\limits_{x\in X}\sum\limits_{a\in A(x)} \alpha_{x,a} = 1;\\ \alpha_{x,a}\ge 0, \ \ \forall x\in X,\ a\in A(x). \end{cases} \tag{7}$$

Based on the mentioned above relationship between problem (3), (4) and problem (5), (7) in [9] the following lemma is proven.

Lemma 1. *Let an average Markov decision problem be given, where an arbitrary stationary strategy s generates a Markov unichain, and consider the function*

$$\psi(s) = \sum_{x\in X}\sum_{a\in A(x)} f(x,a)s_{x,a}\,q_x$$

where q_x for $x\in X$ satisfy the condition

$$\begin{cases} q_y - \sum\limits_{x\in X}\sum\limits_{a\in A(x)} p^a_{x,y}s_{x,a}q_x = 0, \quad \forall y\in X;\\ \sum\limits_{x\in X} q_x = 1. \end{cases}$$

Then the function $\psi(s)$ on the set $\overline{S}$ of solutions of the system

$$\begin{cases} \sum\limits_{a\in A(x)} s_{x,a} = 1, \quad \forall x\in X;\\ s_{x,a}\ge 0, \ \ \forall x\in X,\ a\in A(x) \end{cases}$$

depends only on $s_{x,a}$ for $x\in X, a\in A(x)$, and $\psi(s)$ is quasi-monotone on $\overline{S}$.

Thus, the average unichain decision problem can be represented as the problem of the maximization of a quasi-monotone function $\overline{\psi}(s)$ on a compact set $\overline{S}$. Using this result in [10] it has been shown that an average stochastic game with unichain property can be formulated as a continuous game with quasi-monotone payoffs.

3.2 Determining stationary Nash equilibria for average stochastic games with unichain property

An average stochastic game with unichain property can be formulated in the terms of stationary strategies as follows.

Let $\overline{S} = \overline{S}^1 \times \overline{S}^2 \times \cdots \times \overline{S}^n$, where each $\overline{S}^i$ for $i \in \{1, 2, \ldots, n\}$ represents the set of solutions of system (1), i.e. $\overline{S}^i$ represents the set of mixed stationary strategies for player i. On $\overline{S}$ we define the average payoffs for the players as follows:

$$\psi^i(s^1, s^2, \ \ldots, \ s^n) = \sum_{x \in X} \ \sum_{(a^1, a^2, \ldots, a^n) \in A(x)} \ \prod_{k=1}^{n} \ s^k_{x, a^k} f^i(x, a^1, a^2, \ldots, a^n) q_x,$$

$$i = 1, 2, \ldots, n,$$

where q_x for $x \in X$ are determined uniquely from the following system of linear equations

$$\begin{cases} \displaystyle\sum_{x \in X} \ \sum_{(a^1, a^2, \ldots, a^n) \in A(x)} \ \prod_{k=1}^{n} \ s^k_{x, a^k} p^{(a^1, a^2, \ldots, a^n)}_{x, y} q_x = q_y, \quad \forall y \in X; \\ \displaystyle\sum_{x \in X} q_x = 1, \end{cases}$$

where $s^i \in \overline{S}^i, \ i = 1, 2, \ldots, n.$

The functions $\psi^i(s^1, s^2, \ \ldots, \ s^n), \ i = 1, 2, \ldots, n$ on $\overline{S}$ define a game in normal form that corresponds to a stationary average stochastic game with unichain property. For this game in [11] the following results are proven.

Lemma 2. *For an arbitrary unichain stochastic game each payoff function* $\psi^i(s^1, \ s^2, \ \ldots, \ s^n), \quad i \in \{1, \ 2, \ \ldots, \ n\}$ *possesses the property that* $\psi^i(\overline{s}^1, \overline{s}^2, \ \ldots, \ \overline{s}^{i-1}, s^i, \overline{s}^{i+1}, \ \ldots, \ \overline{s}^n)$ *is quasi-monotone with respect to* $s^i \in \overline{S}^i$ *for arbitrary fixed* $\overline{s}^k \in \overline{S}^k, \ k = 1, 2, \ldots, i-1, i+1, \ldots, n.$

Based on this lemma in [11] the following theorem is proven.

Theorem 1. *Let* $(X, \ A, \ \{X_i\}_{i=\overline{1,n}}, \ \{f^i(x, a)\}_{i=\overline{1,n}}, \ p, \ x)$ *be a stochastic game with a given starting position* $x \in X$ *and average payoff functions*

$$\psi^1(s^1, \, s^2, \ldots, \, s^m), \ \psi^2(s^1, \, s^2, \ldots, \, s^n), \ \ldots, \ \psi^m(s^1, \, s^2, \ldots, \, s^m)$$

of players $1, 2, \ldots, n$, *respectively. If for an arbitrary* $s = (s^1, s^2, \ldots, s^n) \in S$ *of the game the transition probability matrix* $P^s = (p^s_{x,y})$ *corresponds to a Markov unichain then for the continuous game on* $\overline{S}$ *there exists a Nash equilibrium* $s^* = (s^{1^*}, s^{2^*}, \ldots, s^{n*})$ *which is a Nash equilibrium for an arbitrary starting state* $x \in X$ *of the game and* $\psi^i(s^{1^*}, s^{2^*}, \ldots, s^{m*}) = \omega^i_x(s^{1^*}, s^{2^*}, \ldots, s^{m*}), \ \forall x \in X, \ i = 1, 2, \ldots, n.$

4 Some auxiliary results for multichain average decision problem

In this section we propose a continuous model for the multichain average decision problem and extend the results from Section 3.1 for the general case of decision problem. We shall use these results in the next section for the multichain average stochastic games.

4.1 Linear programming approach for multichain decision problem

It is well-known that the optimal stationary strategies for a multichain average Markov decision problem can be found using the following linear programming problem (see [11, 14]):
Maximize

$$\overline{\psi}(\alpha,\beta)=\sum_{x\in X}\sum_{a\in A(x)} f(x,a)\alpha_{x,a} \tag{8}$$

subject to

$$\begin{cases} \sum\limits_{a\in A(y)}\alpha_{y,a}-\sum\limits_{x\in X}\sum\limits_{a\in A(x)}p^a_{x,y}\,\alpha_{x,a}=0, \quad \forall y\in X;\\[2mm] \sum\limits_{a\in A(y)}\alpha_{y,a}+\sum\limits_{a\in A(y)}\beta_{y,a}-\sum\limits_{x\in X}\sum\limits_{a\in A(x)}p^a_{x,y}\beta_{x,a}=\theta_y,\ \ \forall y\in X;\\[2mm] \qquad\alpha_{x,a}\ge 0, \qquad \beta_{y,a}\ge 0, \quad \forall x\in X,\ a\in A(x), \end{cases} \tag{9}$$

where θ_y for $y\in X$ represent arbitrary positive values that satisfy the condition $\sum\limits_{y\in X}\theta_y=1$. Recall that $f(x,a)$ denotes the immediate cost in a state $x\in X$ for a given action $a\in A(x)$ in the decision problem and $p^a_{x,y}$ represent the corresponding probability transitions from a state $x\in X$ to the states $y\in X$ for $a\in A(x)$, where $\sum\limits_{y\in X}p^a_{x,y}=1$.

This problem generalizes the unichain linear programming model (5), (7) from Section 3.1. In (9) the restrictions

$$\sum_{a\in A(y)}\alpha_{y,a}+\sum_{a\in A(y)}\beta_{y,a}-\sum_{x\in X}\sum_{a\in A(x)}p^a_{x,y}\beta_{x,a}=\theta_y, \quad \forall y\in X \tag{10}$$

with the condition $\sum\limits_{y\in X}\theta_y=1$ generalize the constraint

$$\sum_{x\in X}\sum_{a\in A(y)}\alpha_{y,a}=1 \tag{11}$$

in the unichain model. Condition (11) is obtained if we sum (10) over y.

The relationship between feasible solutions of problem (8),(9) and stationary strategies in the average Markov decision problem can be established on the basis of the following *randomized stationary decision rule* (see [14]):

Let (α, β) be a feasible solution of the linear programming problem (8), (9) and denote $X_\alpha = \{x \in X | \sum\limits_{a \in X} \alpha_{x,a} > 0\}$. Then (α, β) possesses the properties that $\sum\limits_{a \in A(x)} \beta_{x,a} > 0$ for $x \in X \setminus X_\alpha$ and a stationary randomized decision rule $d_{\alpha,\beta}(x)$ for a feasible solution (α, β) is defined by

$$s_{d_{\alpha,\beta}(x)}(a) = \begin{cases} \dfrac{\alpha_{x,a}}{\sum\limits_{a \in A(x)} \alpha_{x,a}} & \text{if } x \in X_\alpha; \\ \dfrac{\beta_{x,a}}{\sum\limits_{a \in A(x)} \beta_{x,a}} & \text{if } x \in X \setminus X_\alpha, \end{cases} \tag{12}$$

where $s_{d_{x,y}(x)}(a)$ expresses the probability of choosing the actions $a \in A(x)$ in the states $x \in X$ for the average decision problem under decision rule d. This means that for a given feasible solution (α, β) the decision rule d determines a stationary strategy $s_{x,a} = s_{d_{\alpha,\beta}(x)}(a)$ of choosing the actions $a \in A(x)$ in the states $x \in X$. If for each $x \in X_\alpha$ it holds $\alpha_{x,a} > 0$ for a single $a \in A(x)$ and for each $x \in X \setminus X_\alpha$ it holds $\beta_{x,a} > 0$ for a single $a \in A(x)$ then (12) generates a deterministic decision rule

$$d_{\alpha,\beta}(x) = \begin{cases} a & \text{if } \alpha_{x,a} > 0 \text{ and } x \in X_\alpha; \\ a' & \text{if } \beta_{x,a'} > 0 \text{ and } x \in X \setminus X_\alpha \end{cases}$$

that corresponds to a pure stationary strategy s, where $s_{x,a} = s_{d_{\alpha,\beta}(x)}(a)$ for $x \in X$ and $a \in A(x)$.

Remark 1. In [14] problem (8), (9) is regarded as the dual model of the following linear programming problem:
Minimize

$$\phi(\varepsilon, \omega) = \sum_{x \in X} \theta_x \omega_x \tag{13}$$

subject to

$$\begin{cases} \varepsilon_x + \omega_x \geq f(x,a) + \sum\limits_{y \in X} p^a_{x,y} \varepsilon_y, & \forall x \in X, \ \ \forall a \in A(x); \\ \omega_x \geq \sum\limits_{y \in X} p^a_{x,y} \omega_y, & \forall x \in X, \ \ \forall a \in A(x). \end{cases} \tag{14}$$

The optimal value of objective function in this problem as well as the optimal value of objective function in problem (8), (9) express the optimal average reward when the initial state is chosen according to distribution $\{\theta_x\}$. Solving problem (13), (14) we obtain the value ω^*_x for each $x \in X$ that represents the optimal average reward when transition starts in x with probability equal to 1. This means that if (α^*, β^*) is the optimal solution of problem (8), (9) then we can determine the optimal strategy s^* and the optimal values of object functions of problems (13), (14) and (8), (9), where $\phi(\varepsilon^*, \omega^*) = \overline{\psi}(\alpha^*, \beta^*)$.

4.2 Multichain decision model in the terms of stationary strategies

The continuous model we propose for the multichain average decision problem that generalizes the unichain continuous model (3), (4) is the following:
Maximize
$$\psi(s,q,w)=\sum_{x\in X}\sum_{a\in A(x)} f(x,a)s_{x,a}q_x \tag{15}$$
subject to
$$\begin{cases} q_y-\sum\limits_{x\in X}\sum\limits_{a\in A(x)} p^a_{x,y}\, s_{x,a}q_x=0, & \forall y\in X;\\ q_y+w_y-\sum\limits_{x\in X}\sum\limits_{a\in A(x)} p^a_{x,y}s_{x,a}w_x=\theta_y, & \forall y\in X;\\ \sum\limits_{a\in A(y)} s_{y,a}=1, & \forall y\in X;\\ s_{x,a}\ge 0,\ \ \forall x\in X,\ \forall a\in A(x);\quad w_x\ge 0,\ \forall x\in X, \end{cases} \tag{16}$$
where θ_y are the same values as in problem (8), (9) and $s_{x,a},\ q_x,\ w_x$ for $x\in X$, $a\in A(x)$ represent the variables that must be found.

Theorem 2. *Optimization problem (15), (16) determines the optimal stationary strategies of the multichain average Markov decision problem.*

Proof. Indeed, if we assume that each action set $A(x), x\in X$ contains a single action a' then system (9) is transformed into the following system of equations
$$\begin{cases} q_y-\sum\limits_{x\in X} p_{x,y}q_x=0, & \forall y\in X;\\ q_y+w_y-\sum\limits_{x\in X} p_{x,y}w_x=\theta_y, & \forall y\in X \end{cases}$$
with conditions $q_y,w_y\ge 0$ for $y\in X$, where $q_y=\alpha_{y,a'},\ w_y=\beta_{y,a'},\ \forall y\in X$ and $p_{x,y}=p^{a'}_{x,y},\ \forall x,y\in X$. This system uniquely determines q_x for $x\in X$ and determines w_x for $x\in X$ up to an additive constant in each recurrent class of $P=(p_{x,y})$ (see [14]). Here q_x represents the limiting probability in the state x when transitions start in the states $y\in X$ with probabilities θ_y and therefore the condition $q_x\ge 0$ for $x\in X$ can be released. Note that w_x for some states may be negative, however always the additive constants in the corresponding recurrent classes can be chosen so that w_x became nonnegative. In general, we can observe that in (16) the condition $w_x\ge 0$ for $x\in X$ can be released and this does not influence the value of objective function of the problem. In the case $|A(x)|=1,\ \forall x\in X$ the average cost is determined as $\psi=\sum\limits_{x\in X} f(x)q_x$, where $f(x)=f(x,a), \forall x\in X$.

If the action sets $A(x),\ x\in X$ may contain more than one action then for a given stationary strategy $s\in\overline{S}$ of selection of the actions in the states we can find the average cost $\psi(s)$ in a similar way as above by considering the probability matrix $P^s = (p^s_{x,y})$, where

$$p^s_{x,y} = \sum_{a \in A(x)} p^a_{x,y} s_{x,a} \tag{17}$$

expresses the probability transition from a state $x \in X$ to a state $y \in X$ when the strategy s of selections of the actions in the states is applied. This means that we have to solve the following system of equations

$$\begin{cases} q_y - \sum\limits_{x \in X} p^s_{x,y} q_x = 0, & \forall y \in X; \\ q_y + w_y - \sum\limits_{x \in X} p^s_{x,y} w_x = \theta_y, & \forall y \in X. \end{cases}$$

If in this system we take into account (17) then this system can be written as follows

$$\begin{cases} q_y - \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, s_{x,a} q_x = 0, & \forall y \in X; \\ q_y + w_y - \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y} s_{x,a} w_x = \theta_y, & \forall y \in X. \end{cases} \tag{18}$$

An arbitrary solution (q, w) of the system of equations (18) uniquely determines q_y for $y \in X$ that allows us to determine the average cost per transition

$$\psi(s) = \sum_{x \in X} \sum_{a \in X} f(x, a) s_{x,a} q_x \tag{19}$$

when the stationary strategy s is applied. If we are seeking for an optimal stationary strategy then we should add to (18) the conditions

$$\sum_{a \in A(x)} s_{x,a} = 1, \ \forall x \in X; \ s_{x,a} \geq 0, \ \forall x \in X, a \in A(x) \tag{20}$$

and to maximize (19) under the constraints (18), (20). In such a way we obtain problem (15), (16) without conditions $w_x \geq 0$ for $x \in X$. As we have noted the conditions $w_x \geq 0$ for $x \in X$ do not influence the values of the objective function (15) and therefore we can preserve such conditions that show the relationship of the problem (15), (16) with problem (8), (9). □

The relationship between feasible solutions of problem (8), (9) and feasible solutions of problem (15), (16) can be established on the basis of the following lemma.

Lemma 3. *Let* (s, q, w) *be a feasible solution of problem (15), (16). Then*

$$\alpha_{x,a} = s_{x,a} q_x, \quad \beta_{x,a} = s_{x,a} w_x, \quad \forall x \in X, a \in A(x) \tag{21}$$

represent a feasible solution (α, β) *of problem (8), (9) and* $\varphi(s, q, w) = \overline{\psi}(\alpha, \beta)$. *If* (α, β) *is a feasible solution of problem (8), (9) then a feasible solution* (s, q, w) *of* *problem (15), (16) can be determined as follows:*

$$s_{x,a} = \begin{cases} \dfrac{\alpha_{x,a}}{\sum\limits_{a\in A(x)} \alpha_{x,a}} & \textit{for} \ \ x \in X_\alpha, \ a \in A(x); \\ \dfrac{\beta_{x,a}}{\sum\limits_{a\in A(x)} \beta_{x,a}} & \textit{for} \ \ x \in X \setminus X_\alpha, \ a \in A(x); \end{cases} \tag{22}$$

$$q_x \ = \sum_{a\in A(x)} \alpha_{x,a}, \quad w_x \ = \sum_{a\in A(x)} \beta_{x,a} \quad \textit{for} \ \ x \in X.$$

Proof. Assume that (s, q, w) is a feasible solution of problem (15), (16) and (α, β) is determined according to (21). Then by introducing (21) in (8),(9) we can observe that (9) is transformed in (16) and $\psi(s, q, w) = \overline{\psi}(\alpha, \beta)$, i.e. (α, β) is a feasible solution of problem (8), (9). The second part of lemma follows directly from the properties of feasible solutions of problems (8),(9) and (15),(16). □

Note that an arbitrary pure stationary strategy s of problem (15), (16) corresponds to a basic solution (α, β) of problem (8), (9) for which (22) holds, however system (9) may contain basic solutions for which stationary strategies determined through (22) do not correspond to pure stationary strategies. Moreover two different feasible solutions of problem (8), (9) may generate through (22) the same stationary strategy. Such solutions of system (9) are considered *equivalent solutions* for the decision problem.

Corollary 1. *If* (α^i, β^i), $\ i = \overline{1,k}$, *represent the basic solutions of system (9) then the set of solutions*

$$M = \Big\{(\alpha, \beta) |\ (\alpha, \beta) = \sum_{i=1}^{k} \lambda^i(\alpha^i, \beta^i), \quad \sum_{i=1}^{k} \lambda^i = 1, \quad \lambda^i > 0, \ i = \overline{1,k}\Big\}$$

determines all feasible stationary strategies of problem (15), (16) through (22).

An arbitrary solution (α, β) *of system (9) can be represented as follows:* $\alpha = \sum_{i=1}^{k} \lambda^i \alpha^i$, *where* $\sum_{i=1}^{k} \lambda^i = 1$; $\ \lambda^i \geq 0$, $\ i = \overline{1,k}$, *and* β *represents a solution of the system*

$$\begin{cases} \sum\limits_{a\in A(y)} \beta_{x,a} - \sum\limits_{z\in X} \sum\limits_{a\in A(z)} p^a_{z,x} \beta_{z,a} = \theta_x - \sum\limits_{a\in A(x)} \alpha_{x,a}, \ \forall x \in X; \\ \beta_{y,a} \geq 0, \quad \forall x \in X, \ a \in A(x). \end{cases}$$

If (α, β) *is a feasible solution of problem (8), (9) and* $(\alpha, \beta) \notin M$ *then there exists a solution* $(\alpha', \beta') \in M$ *that is equivalent to* (α, β) *and* $\overline{\psi}(\alpha, \beta) = \overline{\psi}(\alpha', \beta')$.

4.3 The main property of the object function

Using problem (15), (16) we can now extend the results from Section 3.1 for the general case of average decision problem.

Theorem 3. *Let an average Markov decision problem be given and consider the function*

$$\psi(s) = \sum_{x\in X} \sum_{a\in A(x)} f_{(x,a)} s_{x,a}\, q_x, \tag{23}$$

where q_x *for* $x \in X$ *satisfy the condition*

$$\begin{cases} q_y - \sum\limits_{x\in X} \sum\limits_{a\in A(x)} p^a_{x,y}\, s_{x,a} q_x = 0, & \forall y \in X; \\ q_y + w_y - \sum\limits_{x\in X} \sum\limits_{a\in A(x)} p^a_{x,y} s_{x,a} w_x = \theta_y, & \forall y \in X. \end{cases} \tag{24}$$

Then on the set $\overline{S}$ *of solutions of the system*

$$\begin{cases} \sum\limits_{a\in A(x)} s_{x,a} = 1, \quad \forall x \in X; \\ \qquad s_{x,a} \geq 0, \quad \forall x \in X,\ a \in A(x) \end{cases} \tag{25}$$

the function $\psi(s)$ *depends only on* $s_{x,a}$ *for* $x \in X, a \in A(x)$ *and* $\psi(s)$ *is quasi-monotone on* $\overline{S}$ *.*

Proof. For an arbitrary $s \in \overline{S}$ system (24) uniquely determines q_x for $x \in X$ and determines w_x for $x \in X$ up to a constant in each recurrent class of $P^s = (p^s_{x,y})$, where $p^s_{x,y} = \sum\limits_{a\in A(x)} p^a_{x,y} s_{x,a}$, $\forall x, y \in X$. This means that $\psi(s)$ is determined uniquely for an arbitrary $s \in \overline{S}$, i.e. the first part of the theorem holds.

Now let us prove the second part of the theorem.

Consider arbitrary two strategies s', $s'' \in \overline{S}$ and assume that $s' \neq s''$. Then according to Lemma 3 there exist feasible solutions (α', β') and (α'', β'') of linear programming problem (8), (9) for which

$$\psi(s') = \overline{\psi}(\alpha', \beta'), \quad \psi(s'') = \overline{\psi}(\alpha'', \beta'''), \tag{26}$$

where

$$\alpha'_{x,a} = s'_{x,a} q'_x, \quad \alpha''_{x,y} = s''_{x,a} q''_x, \quad \forall x \in X,\ a \in A(x);$$

$$\beta'_{x,a} = s'_{x,a} w'_x, \quad \beta''_{x,y} = s''_{x,a} q''_x, \quad \forall x \in X,\ a \in A(x);$$

$$q'_x = \sum_{a\in A(x)} \alpha'_{x,a} \quad w'_{x,a} = \sum_{a\in A(x)} \beta'_{x,a}, \quad \forall x \in X;$$

$$q''_x = \sum_{a\in A(x)} \alpha''_{x,a} \quad w''_{x,a} = \sum_{a\in A(x)} \beta''_{x,a}, \quad \forall x \in X.$$

The function $\overline{\psi}(\alpha,\beta))$ is linear and therefore for an arbitrary feasible solution $(\overline{\alpha},\overline{\beta})$ of problem (8), (9) holds

$$\overline{\psi}(\overline{\alpha},\overline{\beta}) = t\overline{\psi}(\alpha',\beta') + (1-t)\overline{\psi}(\alpha'',\beta'') \tag{27}$$

if $0 \le t \le 1$ and

$$(\overline{\alpha},\overline{\beta}) = t(\alpha',\beta') + (1-t)(\alpha'',\beta'').$$

Note that $(\overline{\alpha},\overline{\beta})$ corresponds to a stationary strategy $\overline{s}$ for which

$$\psi(\overline{s}) = \overline{\psi}(\overline{\alpha},\overline{\beta}), \tag{28}$$

where

$$\overline{s}_{x,a} = \begin{cases} \dfrac{\overline{\alpha}_{x,a}}{\overline{q}_x} & \text{if } \ x \in X_{\overline{\alpha}}; \\ \dfrac{\beta_{x,a}}{\overline{w}_x} & \text{if } \ x \in X \setminus X_{\overline{\alpha}}. \end{cases} \tag{29}$$

Here $X_{\overline{\alpha}} = \{x \in X |\ \sum\limits_{a\in A(x)} \overline{\alpha}_{x,a} > 0\}$ is the set of recurrent states induced by $P^{\overline{s}} = (p^{\overline{s}}_{x,y})$, where $p^{\overline{s}}_{x,y}$ are calculated according to (17) for $s = \overline{s}$ and

$$\overline{q}_x = tq'_x + (1-t)q'', \quad \overline{w}_x = tw'_x + (1-t)w''_x, \quad \forall x \in X.$$

We can see that $X_{\overline{\alpha}} = X_{\alpha'} \cup X_{\alpha''}$, where $X_{\alpha'} = \{x \in X |\ \sum\limits_{a\in A(x)} \alpha'_{x,a} > 0\}$ and $X_{\alpha''} = \{x \in X |\ \sum\limits_{a\in A(x)} \alpha''_{x,a} > 0\}$.

The value

$$\psi(\overline{s}) = \sum_{x\in X}\sum_{a\in A(x)} f(x,a)\overline{s}_{x,a}\overline{q}_x$$

is determined by $f(x,a)$, $\overline{s}_{x,a}$ and $\overline{q}_x$ in recurrent states $x \in X_{\overline{\alpha}}$ and it is equal to $\overline{\psi}(\overline{\alpha},\overline{\beta})$. If we use (29) then for $x \in X_{\overline{\alpha}}$ and $a \in A(x)$ we have

$$\overline{s}_{x,a} = \frac{t\alpha'_{x,a} + (1-t)\alpha''_{x,a}}{tq'_x + (1-t)q''_x} = \frac{ts'_{x,a}q'_x + (1-t)s''_{x,a}q''_x}{tq'_x + (1-t)q''_x} =$$

$$= \frac{tq'_x}{tq'_x + (1-t)q''_x}s'_{x,a} + \frac{(1-t)q''_x}{tq'_x + (1-t)q''_x}s''_{x,a}$$

and for $x \in X \setminus X_{\overline{\alpha}}$ and $a \in A(x)$ we have

$$\overline{s}_{x,a} = \frac{t\beta'_{x,a} + (1-t)\beta''_{x,a}}{tw'_x + (1-t)w''_x} = \frac{ts'_{x,a}w'_x + (1-t)s''_{x,a}w''_x}{tw'_x + (1-t)w''_x} =$$

$$= \frac{tw'_x}{tw'_x + (1-t)w''_x}s'_{x,a} + \frac{(1-t)w''_x}{tw'_x + (1-t)w''_x}s''_{x,a}.$$

So, we obtain

$$\overline{s}_{x,a} = t_x s'_{x,a} + (1-t_x)s''_{x,a}, \quad \forall a \in A(x), \tag{30}$$

where

$$t_x = \begin{cases} \dfrac{tq'_x}{tq'_x + (1-t)q''_x} & \text{if } \ x \in X_{\overline{\alpha}}; \\ \dfrac{tw'_x}{tw'_x + (1-t)w''_x} & \text{if } \ x \in X \setminus X_{\overline{\alpha}}. \end{cases} \tag{31}$$

and from (26)–(28) we have

$$\psi(\overline{s}) = t\psi(s') + (1-t)\psi(s''). \tag{32}$$

This means that if we consider the set of strategies

$$S(s',s'') = \{\overline{s}|\ \overline{s}_{x,a} = t_x s'_{x,a} + (1-t_x)s''_{x,a}, \quad \forall x \in X, a \in A(x)\}$$

then for an arbitrary $\overline{s} \in S(s',s'')$ holds

$$\min\{\psi(s'), \psi(s'')\} \le \psi(\overline{s}) \le \max\{\psi(s'), \psi(s'')\}, \tag{33}$$

i.e $\psi(s)$ is monotone on $S(s',s'')$. Moreover, using (30)–(33) we obtain that $\overline{s}$ possesses the properties

$$\lim_{t\to 1} \overline{s}_{x,a} = s'_{x,a}, \forall x \in X, a \in A(x); \qquad \lim_{t \to 0} \overline{s}_{x,a} = s''_{x,a}, \forall x \in X, a \in A(x). \tag{34}$$

and respectively

$$\lim_{t\to 1} \psi(\overline{s}) = \psi(s'); \qquad \lim_{t\to 0} \psi(\overline{s}) = \psi(s'').$$

In the following we show that the function $\psi(s)$ is quasi-monotone on $\overline{S}$. To prove this it is sufficient to show that for an arbitrary $c \in R$ the *sublevel set*

$$L^-_c(\psi) = \{s \in \overline{S}|\ \psi(s) \le c\}$$

and the *superlevel set*

$$L^+_c(\psi) = \{s \in \overline{S}|\ \psi(s) \ge c\}$$

of function $\psi(s)$ are convex. These sets can be obtained respectively from the *sublevel set*

$$L^-_c(\overline{\psi}) = \{(\alpha,\beta)|\ \overline{\psi}(\alpha,\beta)) \le c\}$$

and the *superlevel set*

$$L^+_c(\overline{\psi}) = \{(\alpha,\beta)|\ \overline{\psi}(\alpha,\beta)) \ge c\}$$

of function $\overline{\psi}(\alpha,\beta)$ for linear programming problem (8), (9) using (22).

Denote by $(\alpha^i, \beta^i), \quad i = \overline{1,k}$ the basic solutions of system (9). According to Corollary 1 all feasible strategies of problem (8), (9) can be obtained trough (22) using the basic solutions (α^i, β^i), $\quad i = \overline{1,k}$. Each (α^i, β^i), $\quad i = \overline{1,k}$, determines a stationary strategy

$$s^i_{x,a} = \begin{cases} \dfrac{\alpha^i_{x,a}}{q^i_x}, & \text{for } \; x \in X_{\alpha^i}, \; a \in A(x); \\ \dfrac{\beta^i_{x,a}}{w^i_x}, & \text{for } \; x \in X \setminus X_{\alpha^i}, \; a \in A(x) \end{cases} \tag{35}$$

for which $\psi(s^i) = \overline{\psi}(\alpha^i, \beta^i)$ where

$$X_{\alpha^i} = \{x \in X | \sum_{a \in A(x)} \alpha^i_{x,a} > 0\}, \quad q^i_x = \sum_{a \in A(x)} \alpha^i_{x,a}, \quad w^i_x = \sum_{a \in A(x)} \beta^i_{x,a}, \; \forall x \in X. \tag{36}$$

An arbitrary feasible solution (α, β) of system (9) determines a stationary strategy

$$s_{x,a} = \begin{cases} \dfrac{\alpha_{x,a}}{q_x}, & \text{for } \; x \in X_{\alpha}, \; a \in A(x); \\ \dfrac{\beta_{x,a}}{w_x}, & \text{for } \; x \in X \setminus X_{\alpha}, \; a \in A(x), \end{cases} \tag{37}$$

for which $\psi(s) = \overline{\psi}(\alpha, \beta)$ where

$$X_{\alpha} = \{x \in X | \sum_{a \in A(x)} \alpha_{x,a} > 0\}, \qquad q_x = \sum_{a \in A(x)} \alpha_{x,a}, \quad w_x = \sum_{a \in A(x)} \beta_{x,a}, \; \forall x \in X.$$

Taking into account that (α, β) can be represented as

$$(\alpha, \beta) = \sum_{i=1}^{k} \lambda^i (\alpha^i, \beta^i), \; \text{where} \; \sum_{i=1}^{k} \lambda^i = 1, \quad \lambda^i \geq 0, \; i = \overline{1,k} \tag{38}$$

we have $\overline{\psi}(\alpha, \beta) = \sum\limits_{i=1}^{k} \overline{\psi}(\alpha^i, \beta^i)\lambda^i$ and we can consider

$$X_{\alpha} = \bigcup_{i=1}^{k} X_{\alpha^i}; \quad \alpha = \sum_{i=1}^{k} \lambda^i \alpha^i; \quad q = \sum_{i=1}^{k} \lambda^i q^i; \quad w = \sum_{i=1}^{k} \lambda^i w^i. \tag{39}$$

Using (35)–(39) we obtain:

$$s_{x,a} = \frac{\alpha_{x,a}}{q_x} = \frac{\sum_{i=1}^{k} \lambda^i \alpha^k_{x,a}}{q_x} = \frac{\sum_{i=1}^{k} \lambda^i s^i_{x,a} q^i_x}{q_x} = \sum_{i=1}^{k} \frac{\lambda^i q^i_x}{q_x} s^i_{x,a}, \quad \forall x \in X_{\alpha}, \; a \in A(x);$$

$$s_{x,a} = \frac{\beta_{x,a}}{w_x} = \frac{\sum_{i=1}^{k} \lambda^i \beta^k_{x,a}}{w_x} = \frac{\sum_{i=1}^{k} \lambda^i s^i_{x,a} w^i_x}{w_x} = \sum_{i=1}^{k} \frac{\lambda^i w^i_x}{w_x} s^i_{x,a}, \; \forall x \in X \backslash X_{\alpha}, a \in A(x)$$

and

$$q_x = \sum_{i=1}^{k} \lambda^i q^i_x, \qquad w_x = \sum_{i=1}^{k} \lambda^i w^i_x \; \text{ for } \quad x \in X. \tag{40}$$

So,

$$s_{x,a} = \begin{cases} \displaystyle\sum_{i=1}^{k} \frac{\lambda^i q_x^i}{q_x} s_{x,a}^i & \text{if } \ q_x > 0; \\ \displaystyle\sum_{i=1}^{k} \frac{\lambda^i w_x^i}{w_x} s_{x,a}^i & \text{if } \ q_x = 0, \end{cases} \tag{41}$$

where q_x and w_x are determined according to (40).

We can see that if $\lambda^i, s^i, q^i, \ i = \overline{1,k}$ are given then the strategy s defined by (41) is a feasible strategy because $s_{x,a} \geq 0, \forall x \in X, a \in A(x)$ and $\sum_{a\in A(x)} s_{x,a} = 1, \ \forall x \in X$. Moreover, we can observe that $q_x = \sum\limits_{i=1}^{k} \lambda^i q_x^i, \quad w_x = \sum\limits_{i=1}^{k} \lambda^i w_x^i$ for $x \in X$ represent a solution of system (24) for the strategy s defined by (41). This can be verified by introducing (40) and (41) in (24); after such a substitution all equations from (24) are transformed into identities. For $\psi(s)$ we have

$$\psi(s) = \sum_{x\in X}\sum_{a\in A(x)} f(x,a)s_{x,a}q_x = \sum_{x\in X_\alpha}\sum_{a\in A(x)} f(x,a)\sum_{i=1}^{k}\left(\frac{\lambda^i q_x^i}{q_x} s_{x,a}^i\right)q_x =$$

$$\sum_{i=1}^{k}\Bigg(\sum_{x\in X_{\alpha^i}}\sum_{a\in A(x)} f(x,a)s_{x,a}^i q_x^i\Bigg)\lambda^i = \sum_{i=1}^{k}\psi(s^i)\lambda^i,$$

i.e.

$$\psi(s) = \sum_{i=1}^{k}\psi(s^i)\lambda^i, \tag{42}$$

where s is the strategy that corresponds to (α, β).

Thus, assuming that the strategies $s^1, s^2, \ldots, s^k$ correspond to basic solutions $(\alpha^1, \beta^1), (\alpha^2, \beta^2), \ldots, (\alpha^k, \beta^k)$ of problem (8), (9) and $s \in \overline{S}$ corresponds to an arbitrary solution (α, β) of this problem that can be expressed as convex combination of basic solutions of problem (8), (9) with the corresponding coefficients $\lambda^1, \lambda^2, \ldots, \lambda^k$, we can express the strategy s and the corresponding value $\psi(s)$ by (40)–(42). In general the representation (40)–(42) of strategy s and of the value $\psi(s)$ is valid for an arbitrary finite set of strategies from $\overline{S}$ if (α, β) can be represented as convex combination of the finite number of feasible solutions $(\alpha^1, \beta^1), (\alpha^2, \beta^2), \ldots, (\alpha^k, \beta^k)$ that correspond to $s^1, s^2, \ldots, s^k$; in the case $k = 2$ from (40)–(42) we obtain (30)–(32). It is evident that for a feasible strategy $s \in S$ the representation (40), (41) may be not unique, i.e. two differen vectors $\overline{\Lambda} = (\overline{\lambda}^1, \overline{\lambda}^2, \ldots, \overline{\lambda}^k)$ and $\overline{\overline{\Lambda}} = \overline{\overline{\lambda}}^1, \overline{\overline{\lambda}}^2, \ldots, \overline{\overline{\lambda}}^k$ may be that determine the same strategy s via (40), (41). In the following we will assume that $s^1, s^2, \ldots, s^k$ represent the system of linear independent basic solutions of system (25), i.e. each $s^i \in \overline{S}$ corresponds to a pure stationary strategy.

Thus, an arbitrary strategy $s \in \overline{S}$ is determined according to (40), (41) where $\lambda^1, \lambda^2, \ldots, \lambda^k$ correspond to a solution of the following system

$$\sum_{i=1}^{k} \lambda^i = 1; \;\; \lambda^i \geq 0, \;\; i = \overline{1,k}.$$

Consequently, the sublevel set $L_c^-(\psi)$ of function $\psi(s)$ represents the set of strategies s determined by (40), (41), where $\lambda^1, \lambda^2, \ldots, \lambda^k$ satisfy the condition

$$\begin{cases} \sum\limits_{i=1}^{k} \psi(s^i)\lambda^i \leq c; \\ \sum\limits_{i=1}^{k} \lambda^i = 1; \quad \lambda^i \geq 0, \quad i = \overline{1,k} \end{cases} \tag{43}$$

and the superlevel set $L_c^+(\psi)$ of $\psi(s)$ represents the set of strategies s determined by (40),(41), where $\lambda^1, \lambda^2, \ldots, \lambda^k$ satisfy the condition

$$\begin{cases} \sum\limits_{i=1}^{k} \psi(s^i)\lambda^i \geq c; \\ \sum\limits_{i=1}^{k} \lambda^i = 1; \quad \lambda^i \geq 0, \quad i = \overline{1,k}. \end{cases} \tag{44}$$

Respectively the level set $L_c(\psi) = \{s \in \overline{S} |\ \psi(s) = c\}$ of function $\psi(s)$ represents the set of strategies s determined by (40), (41), where $\lambda^1, \lambda^2, \ldots, \lambda^k$ satisfy the condition

$$\begin{cases} \sum\limits_{i=1}^{k} \psi(s^i)\lambda^i = c; \\ \sum\limits_{i=1}^{k} \lambda^i = 1; \quad \lambda^i \geq 0, \quad i = \overline{1,k}. \end{cases} \tag{45}$$

Let us show that $L_c^-(\psi)$, $L_c^+(\psi)$, $L_c(\psi)$ are convex sets. We present the proof of convexity of sublevel set $L_c^-(\psi)$. The proof of convexity of $L_c^+(\psi)$ and $L_c(\psi)$ is similar to the proof of convexity of $L_c^-(\psi)$.

Denote by Λ the set of solutions $(\lambda^1, \lambda^2, \ldots, \lambda^k)$ of system (43). Then from (40), (41), (43) we have

$$L_c^-(\psi) = \prod_{x \in X} \hat{S}_x$$

where $\hat{S}_x$ represents the set of strategies

$$s_{x,a} = \begin{cases} \dfrac{\sum_{i=1}^{k} \lambda^i q_x^i s_{x,a}^i}{\sum_{i=1}^{k} \lambda^i q_x^i} & \text{if } \sum_{i=1}^{k} \lambda^i q_x^i > 0, \\[2ex] \dfrac{\sum_{i=1}^{k} \lambda^i w_x^i s_{x,a}^i}{\sum_{i=1}^{k} \lambda^i w_x^i} & \text{if } \sum_{i=1}^{k} \lambda^i q_x^i = 0, \end{cases} \qquad a \in A(x)$$

in the state $x \in X$ determined by $(\lambda^1, \lambda^2, \ldots, \lambda^k) \in \Lambda$.

For an arbitrary $x \in X$ the set Λ can be represented as follows $\Lambda = \Lambda_x^+ \cup \Lambda_x^0$, where

$$\Lambda_x^+ = \{(\lambda^1, \lambda^2, \ldots, \lambda^k) \in \Lambda | \sum_{i=1}^{k} \lambda^i q_x^i > 0\},$$

$$\Lambda_x^0 = \{(\lambda^1, \lambda^2, \ldots, \lambda^k) \in \Lambda | \sum_{i=1}^{k} \lambda^i q_x^i = 0\}$$

and $\sum_{i=1}^{k} \lambda^i w_x^i > 0$ if $\sum_{i=1}^{k} \lambda^i q_x^i = 0$.

Therefore $\hat{S}_x$ can be expressed as follows $\hat{S}_x = \hat{S}_x^+ \cup \hat{S}_x^0$, where $\hat{S}_x^+$ represents the set of strategies

$$s_{x,a} = \frac{\sum_{i=1}^{k} \lambda^i q_x^i s_{x,a}^i}{\sum_{i=1}^{k} \lambda^i q_x^i}, \text{ for } a \in A(x) \tag{46}$$

in the state $x \in X$ determined by $(\lambda^1, \lambda^2, \ldots, \lambda^k) \in \Lambda_x^+$ and $\hat{S}_x^0$ represents the set of strategies

$$s_{x,a} = \frac{\sum_{i=1}^{k} \lambda^i w_x^i s_{x,a}^i}{\sum_{i=1}^{k} \lambda^i w_x^i}, \text{ for } a \in A(x) \tag{47}$$

in the state $x \in X$ determined by $(\lambda^1, \lambda^2, \ldots, \lambda^k) \in \Lambda_x^0$.

Therefore $\hat{S}_x$ can be expressed as follows $\hat{S}_x = \hat{S}_x^+ \cup \hat{S}_x^0$, where $\hat{S}_x^+$ represents the set of strategies

$$s_{x,a} = \frac{\sum_{i=1}^{k} \lambda^i q_x^i s_{x,a}^i}{\sum_{i=1}^{k} \lambda^i q_x^i}, \text{ for } a \in A(x) \tag{48}$$

in the state $x \in X$ determined by $(\lambda^1, \lambda^2, \ldots, \lambda^k) \in \Lambda_x^+$ and $\hat{S}_x^0$ represents the set of strategies

$$s_{x,a} = \frac{\sum_{i=1}^{k} \lambda^i w_x^i s_{x,a}^i}{\sum_{i=1}^{k} \lambda^i w_x^i}, \text{ for } a \in A(x) \tag{49}$$

in the state $x \in X$ determined by $(\lambda^1, \lambda^2, \ldots, \lambda^k) \in \Lambda_x^0$.

Thus, if we analyze (48) then observe that $s_{x,a}$ for a given $x \in X$ represents a linear-fractional function with respect to $\lambda^1, \lambda^2, \ldots, \lambda^k$ defined on convex set Λ_x^+ and $\hat{S}_x^+$ is the image of $s_{x,a}$ on Λ_x^+. Therefore $\hat{S}_x^+$ is a convex set. If we analyze (49) then observe that $s_{x,a}$ for given $x \in X$ represents a linear-fractional function with respect to $\lambda^1, \lambda^2, \ldots, \lambda^k$ on convex set Λ_x^0 and $\hat{S}_x^0$ is the image of $s_{x,a}$ on Λ_x^0. Therefore $\hat{S}_x^0$ is a convex set (see [1]). Additionally we can observe that $\Lambda_x^+ \cap \Lambda_x^0 = \emptyset$ and in the case $\Lambda_x^+, \Lambda_x^0, \neq \emptyset$ the set Λ_x^0 represents the limit inferior of Λ_x^+. Using this property and taking into account (34) we can conclude that each strategy $s_x \in \hat{S}_x^0$ can be regarded as the limit of a sequence of strategies $\{s_x^t\}$ from $\hat{S}_x^+$. Therefore we obtain that $\hat{S}_x = \hat{S}_x^+ \cup \hat{S}_x^0$ is a convex set. This involves the convexity of the sublevel set $L_c^-(\psi)$. In analogues way using (44) and (45) we can show that the superlevel set $L_c^+(\psi)$ and the level set $L_c(\psi)$ a convex set. This means that the function $\psi(s)$ is quasi-monotone on $\overline{S}$. □

5 Existence of stationary Nash equilibria for the multichain average stochastic game

In this section we present an result concerned with the existence of stationary Nash equilibria in a multichain average stochastic game with n players. We prove this result using a continuous model for the considered game that generalizes the continuous model from Section 3.

5.1 A continuous model for the multichain stochastic game

The continuous model for a multichain average stochastic game that generalizes the continuous model (23)–(25) is the following:

Let $\overline{S}^i,\ i \in \{1, 2, \dots n\}$ be the set of solutions of the system

$$\begin{cases} \sum\limits_{a_i \in A^i(x)} s^i_{x,a^i} = 1, \quad \forall x \in X; \\ \qquad\quad s^i_{x,a^i} \geq 0, \quad \forall x \in X,\ a^i \in A^i(x). \end{cases} \tag{50}$$

that determines the set of stationary strategies of player i. Each $\overline{S}^i$ is a convex compact set and an arbitrary its extreme point corresponds to a basic solution s^i of system (50), where $s^i_{x,a^i} \in \{0, 1\},\ \forall x \in X,\ a^i \in A(x)$, i.e each basic solution of this system corresponds to a pure stationary strategy of player i. On the set $\overline{S} = \overline{S}^1 \times \overline{S}^2 \times \cdots \times \overline{S}^n$ we define n payoff functions

$$\begin{cases} \psi^i(s^1, s^2,\ \dots,\ s^n) = \sum\limits_{x\in X} \ \sum\limits_{(a^1,a^2,\dots,a^n)\in A(x)} \ \prod\limits_{k=1}^{n} \ s^k_{x,a^k} f^i(x, a^1, a^2 \dots a^n) q_x, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, 2, \dots, n, \end{cases} \tag{51}$$

where q_x for $x \in X$ are determined uniquely from the following system of linear equations

$$\begin{cases} q_y - \sum\limits_{x\in X} \ \sum\limits_{(a^1,a^2,\dots,a^n)\in A(x)} \ \prod\limits_{k=1}^{n} \ s^k_{x,a^k} p^{(a^1,a^2,\dots,a^n)}_{x,y} q_x = 0, \quad \forall y \in X; \\ q_y + w_y - \sum\limits_{x\in X} \ \sum\limits_{(a^1,a^2,\dots,a^n)\in A(x)} \ \prod\limits_{k=1}^{n} \ s^k_{x,a^k} p^{(a^1,a^2,\dots,a^n)}_{x,y} w_x = \theta_x, \forall y \in X, \end{cases} \tag{52}$$

for an arbitrary profile $(s^1, s^2,\ \dots,\ s^m) \in \overline{S}$. Each $(s^1, s^2,\ \dots,\ s^n) \in \overline{S}$ in the considered continuous game corresponds to a profile of mixed stationary strategies of the players and $\psi^i(s^1, s^2,\ \dots,\ s^n),\ i = 1, 2, \dots, n$, defined by (51), (52) represent the corresponding average payoffs of the players in the case when the staring state is chosen according to distribution $\{\theta_x\}$. If $\theta_x = 0,\ \forall x \in X \setminus \{x_0\}$ and $\theta_{x_0} = 1$ then we obtain the continuous game model for the average stochastic game with given starting state x_0, i.e. $\psi^i(s^1, s^2, \dots, s^n) = \omega^i_{x_0}(s^1, s^2, \dots, s^n),\ i = 1, 2, \dots, n.$

5.2 The main result

From Theorem 3 as a corollary we can obtain the following lemma.

Lemma 4. *For an arbitrary average stochastic game with* $\theta_x > 0, \forall x \in X$ *each payoff function* $\psi^i(s^1,\ s^2,\ \dots,\ s^n)$, $i \in \{1,2,\dots,n\}$ *possesses the property that* $\psi^i(\overline{s}^1, \overline{s}^2,\ \dots,\ \overline{s}^{i-1}, s^i, \overline{s}^{i+1},\ \dots,\ \overline{s}^n)$ *is quasi-monotone with respect to* $s^i \in \overline{S}^i$ *for arbitrary fixed* $\overline{s}^k \in \overline{S}^k$, $k = 1,2,\dots,i-1,i+1,\ \dots,\ n$.

Proof. Indeed, if players $1,2,\dots,i-1,i+1,\ \dots,\ n$ fix their stationary strategies $\overline{s}^k \in \overline{S}^k$, $k = 1,2,\dots,i-1,i+1,\ \dots,\ n,$ then we obtain an average decision problem with respect to $s^i \in \overline{S}^i$ and average cost function $\psi^i(\overline{s}^1, \overline{s}^2,\ \dots,\ \overline{s}^{i-1}, s^i, \overline{s}^{i+1},\ \dots,\ \overline{s}^n)$. According to Theorem 3 this function possesses the property that the value of the function is uniquely determined by $s^i \in \overline{S}^i$ and it is quasi-monotone with respect to s^i on $\overline{S}^i$. □

Theorem 4. *Let* $(X, A, \{X_i\}_{i=\overline{1,n}}, \{f^i(x,a)\}_{i=\overline{1,n}}, p, \{\theta_x\})$ *be an average stochastic game with given distribution* $\{\theta_x\}$ *for the initial state and consider the continuous game with average payoffs* $\psi^i(s^1, s^2, \dots, s^n)$, $i = 1,2,\dots,n$ *for the players. If for an arbitrary profile* $\overline{s} = (\overline{s}^1, \overline{s}^2, \dots, \overline{s}^n) \in \overline{S}$ *each payoff function* $\psi^i(s^1, s^2, \dots, s^n)$, $i \in \{1,2,\dots,n\}$ *possesses the property that*

$$\lim_{s^i \to \overline{s}^i} \psi^i(\overline{s}^1, \overline{s}^2, \dots, \overline{s}^{i-1}, s^i, \overline{s}^{i+1}, \dots, \overline{s}^n) = \psi^i(\overline{s}^1, \overline{s}^2, \dots, \overline{s}^{i-1}, \overline{s}^i, \overline{s}^{i+1}, \dots, \overline{s}^n)$$

then for the considered continuous game there exists a Nash equilibrium $s^* = (s^{1^*},\ s^{2^*},\ \dots,\ s^{n*}) \in \overline{S}$ *that is a stationary Nash equilibrium for the average stochastic game* $(X,\ A,\ \{X_i\}_{i=\overline{1,n}},\ \{f^i(x,a)\}_{i=\overline{1,n}},\ p,\ x\})$ *with an arbitrary initial state* $x \in X$.

Proof. According to Lemma 4 each function $\psi^i(s^1, s^2, \dots, s^n)$, $i \in \{1,2,\dots,n\}$ satisfies the condition that $\psi^i(\overline{s}^1, \overline{s}^2,\ \dots,\ \overline{s}^{i-1}, s^i, \overline{s}^{i+1},\ \dots,\ \overline{s}^n)$ is quasi-monotone with respect to $s^i \in \overline{S}^i$ for an arbitrary fixed $\overline{s}^k \in \overline{S}^k$, $k = 1,\ 2,\ \dots,\ i-1,$ $i+1,\ \dots,\ n$. In the considered game each subset $\overline{S}^i$ is convex and compact and according to the conditions of the theorem each payoff function $\psi^i(s^1, s^2, \dots, s^n)$ is continue with respect to s^i in $\overline{S}^i$. Therefore, these conditions (see [2, 3, 15, 18]) provide the existence of a Nash equilibrium $s^* = (s^{1^*}, s^{2^*}, \dots, s^{n*})$ for the game with payoff functions $\psi^i(s^1, s^2,\ \dots,\ s^n)$, $i \in \{i, 2,\ \dots,\ n\}$ on $\overline{S}^1 \times \overline{S}^2 \times \cdots \times \overline{S}^n$. □

6 Conclusion

The results presented in the paper show that for finite state space stochastic games with average payoffs stationary Nash equilibria exist if the conditions of Theorem 4 are satisfied. For determining stationary Nash equilibria in the considered games the continuous model from Section 5.1 can be used. For average stochastic games with unichain property the continuous model from Section 3.2 can be used.

References

[1] Boyd S., Vandenberghe L. *Convex Optimization.* University Press, Cambridge 2004.

[2] Dasgupta P., Maskin E. *The existence of Equilibrium in Discontinuous Economic Games.* Review of Economic Studies, 1986, **53**, 1–26.

[3] Debreu G. *A Social Equilibrium Existence Theorem.* Proceedings of the National Academy of Sciences, 1952, 386–393.

[4] Filar J. A., Vrieze K. *Competitive Markov Decision Processes.* Springer, 1997.

[5] Filar J. A., Schultz T. A., Thuijsman F., Vrieze O. J. *Nonlinear programming and stationary equilibria of stochastic games.* Mathematical Programming, 1991, **5**, 227–237.

[6] Fink A. M. *Equilibrium in a stochastic n-person game.* J. Sci. Hiroshima Univ., Series A-1, 1964, **28**, 89–93.

[7] Flesch J., Thuijsman F., Vrieze K. *Cyclic Markov equilibria in stochastic games*, International Journal of Game Theory, 1977, **26**, 303–3014 66.

[8] Gillette D. *Stochastic games with zero stop probabilities.* Contribution to the Theory of Games, Princeton, 1957, **III**, 179–187.

[9] Lozovanu D. *The game-theoretical approach to Markov decision problems and determining Nash equilibria for stochastic positional games.* Int. J. Mathematical Modelling and Numerical Optimization, 2011, **2(2)**, 162–164.

[10] Lozovanu D., Pickl S. *On Nash equilibria for stochastic games and determining the optimal strategies of the players.* Contribution to game theory and management, St. Petersburg University, 2015, **VIII**, 187–198.

[11] Lozovanu D., Pickl S. *Optimization of Stochastic Discrete Systems and Control on Complex Networks*, Springer, 2015.

[12] Mertens J. F., Neyman A. *Stochastic games.* Int. J. of Game Theory, 1981, **10**, 53–66.

[13] Neyman A., Sorin S. *Stochastic games and applications*, NATO ASI series, Kluver Academic press, 2003.

[14] Puterman M. *Markov Decision Processes: Stochastic Dynamic Programming.* John Wiley, New Jersey, 2005.

[15] Reny F. *On the existence of Pure and Mixed Strategy Nash Equilibria In Discontinuous Games.* Econometrica, 1999, **67**, 1029–1056.

[16] Rogers P. D. *Nonzero-Sum Stochastic games.* PhD thesis, Report ORC 68-8, 1969.

[17] Shapley L. *Stochastic games*, Proc. Natl. Acad. Sci. U.S.A., 1953, **39**, 1095–1100.

[18] Simon L. *Games with Discontinuous Payoffs.* Review of Economic Studies, 1987, **54**, 569–597.

[19] Sobel M. *Noncooperative stochastic games.* The Annals of Mathematical statistics, 1971, **42**, 1930–1035.

[20] Vieille N. *Equilibrium in 2-person stochastic games I, II.* Israel J. Math. 2000, **119(1)**, 55–126.

[21] Vrieze O. V. *Stochastic games with finite state and actions spaces.* CWI-Tract, Center of Mathematics and Computer Science, Amsterdam, 1987, **33**, 295–320.

Dmitrii Lozovanu
Institute of Mathematics and Computer Science
5 Academiei str., Chişinău, MD−2028
Moldova

E-mail: *lozovanu@math.md*

Received April 6, 2016