STUDIA UNIVERSITATIS

Revistă științifică a Universității de Stat din Moldova, 2010, nr.7(37)

LINEAR PROGRAMMING APPROACH FOR MARKOV DECISION PROBLEMS WITH AVERAGE COST OPTIMIZATION CRITERION

Maria CAPCELEA

Chair of Applied Mathematics

În lucrare este demonstrată echivalența problemei decizionale Markov cu problema stocastică de control optimal pe rețea. Acest rezultat permite de a aplica metoda programării liniare la determinarea strategiilor staționare optimale în procesele decizionale Markov cu criteriul de optimizare a costului mediu.

1. Introduction and Problem Formulation

The linear programming approach we extend for Markov decision problem with average cost optimization criterion. We show that an arbitrary Markov decision problem can be transformed into a stochastic control problem on network, and vice versa an arbitrary stochastic control problem on network can be formulated as a Markov decision problem. Thus, the considered problems are equivalent and therefore the linear programming approach can be developed and specified for Markov decision problems.

A Markov decision process [1-3] is determined by a tuple (X, A, p, c), where X is a finite state space, A is a finite set of actions, p is a nonnegative real function $p: A\times X \times X \to R^+$ that satisfy the condition $\sum_{y\in X} p^a_{x,y} = 1,\ \forall a \in A$ and c is an arbitrary real function $c: A\times X \times X \to \mathbb{R}$.

The function p for a fixed action $a \in A$ and arbitrary $x, y \in X$ determines the probability $p^a_{x,y}$ of the system's transition from the state $x \in X$ at the moment of time t to state y at the moment of time $t+1$ for every $t = 0,1,2,\ldots$. The function c for a fixed action $a \in X$ and arbitrary $x, y \in X$ determine the cost $c^a_{x,y}$ of system's transition from the state x to the state y when the system makes transition from x to y with the probability $p^a_{x,y}$. In the considered Markov process the functions p and c do not depend on time, i.e. we have a stationary Markov decision process. We assume that at the moment of time $t = 0$ the dynamical system is in the state x_{i_0}.

A stationary strategy s in the Markov process we define as a map

$$s: x \to a \in A(x) \quad for \quad x \in X,$$

where $A(x)$ represents the set of actions in the state $x \in X$. An arbitrary stationary strategy s induces a simple Markov process with the transition probability matrix $P^s = (p^s_{x,y})$ and transition cost matrix $C^s = (c_{x,y})$. For this Markov process we can determine the expected average cost per transition $\omega^s_{x_{i_0}}$ when the dynamical system starts transitions in the state x_{i_0} at the moment of time $t = 0$. This quantity we denote $f_{x_{i_0}}(s)$, i.e.

$$f_{x_{i_0}}(s) = \omega^s_{x_{i_0}}.$$

We consider the Markov decision problem with average cost criterion, i.e we are seeking for a strategy s^* for which

$$f_{x_{i_0}}(s^*) = \min_s f_{x_{i_0}}(s).$$

For an arbitrary Markov decision problem we may assume that the action sets in different states are different, i.e. $A(x) \neq A(y)$. However it is easy to observe that an arbitrary problem can be reduced to the case $|A(x)| = |A(y)| = |A|, \forall x, y \in X$ introducing some copies of the actions in the states $y \in X$ if for two different states $x, y \in X$ holds $|A(y)| < |A(x)|$.

In the case $|A(x)|=|A(y)|=|A|, \forall x, y \in X$ a Markov decision process can be given by $2|A|$ matrices $P^{a_k}=(p_{x,y}^{a_k}), C^{a_k}=(c_{x,y}^{a}), k=1,2,\ldots,|A|$, where $\sum_{y\in X} p_{x,y}^{a_k}=1, \forall a_k \in A$. A fixed strategy

$$s: x \to a_k \in A(x) \text{ for } x \in X$$

generates a simple Markov process with the probability transition matrix P^s and transition cost matrix C^s induced by the roes of the corresponding matrices P^{a_k} and C^{a_k}, $k=1,2,\ldots,|A|$, respectively.

Using the matrix representation of the Markov decision processes we can show that the stochastic control problem with average cost criterion can be represented as a Markov decision problem. Indeed, the matrix representation of the control problem corresponds to the case when $X = X_1 \cup X_2$ $(X_1 \cap X_2)$, where for an arbitrary state $x_i \in X_1$ the probabilities $p_{x_i,y}^{a_k}$ are equal to 0 or 1 and for an arbitrary state $x_i \in X_2$ the corresponding $i-th$ rows in the matrices $P^{a_1}, P^{a_2}, \ldots, P^{a_{|A|}}$ and $C^{a_1}, C^{a_2}, \ldots, C^{a_{|A|}}$ are the same. This means that an arbitrary stochastic control represent a particulary case of the Markov decision problem.

In the next section we show that an arbitrary Markov decision problem with average cost criterion can be reduced to a stochastic control problem on an auxiliary network. In a such way we prove that the considered problems are equivalent. Using the reduction procedure of Markov decision problem to stochastic control problem we propose an algorithm for determining the optimal stationary strategies for Markov decision problem.

2. Algorithm for Solving Markov Decision Problem Using a Reduction Procedure to Stochastic Control Problem

Let us show that the problem of determining the optimal stationary strategies s^* in a Markov decision process (X, A, p, c) with average cost can be reduced to the problem of determining the optimal stationary strategy in the control problem on a network $(G', X_1', X_2', p', c', x_{i_0}')$, where $G'=(X', E')$, X_1', X_2', p', c' and x_{i_0} are defined in the following way. The set of vertices $X' = X_1' \cup X_2'$ contains $(|A|+1)|X|$ vertices, where $|X_1'|=|X|$ and $|X_2'|=|A||X|$. So, the set of controllable states in the control problem consists of a copy of set of states X and the set of uncontrollable states X_2' consists of $|A|$ copies of the set of states X. Strictly X_1' and X_2' we define as follows:

$$X_1' = \{x' = x \mid x \in X\}; \quad X_2' = \bigcup_{a\in A} X^a,$$

where

$$X^a = \{x^a = (x,a) \mid x \in X\}, \ \forall a \in A.$$

The set of directed edges E' we also represent as a couple of two disjoint subsets $E' = E_1' \cup E_2'$, where E_1' is the set of outgoing edges from $x' \in X_1$ and E_2' is the set of outgoing edges from $x^a \in X_2'$. The sets E_1' and E_2' are defined as follow:

$$E_1' = \{(x,(x,a)) \mid x \in X_1; \ (x,a) \in X_2, \ a \in A\};$$

$$E_2' = \{((x,a),y) \mid (x,a) \in X_2', \ y \in X_1, \ p_{x,y}^a > 0, \ a \in A\}.$$

On the set of directed edges E' we define the cost function $c': E' \to \mathbb{R}$, where

$$c_{e'}' = 0, \ \forall e' = (x,(x,a)) \in E_1'; \ c_{e'}' = c_{x,y}^a \text{ for } e' = ((x,a),y) \in E_2' \quad (x, y \in X, a \in A).$$

On E_2' we define the transition probability function $p': E_2' \to [0,1]$, where $p_{e'}' = p_{x,y}^a$ *for* $e' = ((x,a),y) \in E_2'$.

It is easy to observe that between the set of stationary strategies S in the Markov decision process and the set of strategies S' in the control problem on network $(G', X_1, X_2, p', c', x_{i_0})$ there exist a bijective mapping that preserve the average cost per transition. Therefore if we find the optimal stationary strategy for the control problem on network then we can determined the optimal stationary strategy in Markov decision process.

The network constructed above gives a graphical interpretation of the Markov decision process via the structure of the graph G, where the actions and all possible transitions for an arbitrary fixed action are represented by arcs and nodes. A more simple graphical interpretation of the Markov decision process may be given by the multigraph $GM=(X,EM)$ (graph with parallel directed edges) with the set of vertices X that corresponds to the set of states and the set of edges EM that consists of $|A|$ subsets $EM_1, EM_2, \ldots, EM_{|A|}$ $\left(EM=\bigcup_{i=1}^{|A|}EM_i\right)$, where $EM_i=\{e^{a_i}=(x,y)^{a_i} \mid p_{x,y}^{a_i}>>0\}$, $i=1,2,\ldots,|A(x)|$.

The graphical interpretation of the Markov decision process and an example how to solve the decision problem using reduction procedure to an auxiliary control problem on network are given bellow.

Example. Consider a Markov decision process (X,A,p,c) where $X=\{1,2\}$, $A=1,2$ and the possible values of the corresponding probability and cost functions $p: X\times X\times A\to[0,1]$, $c: X\times X\times A\to\mathbb{R}$ are defined as follows:

$$p_{1,1}^{a_1}=0.7,\ p_{1,2}^{a_1}=0.3,\ p_{2,1}^{a_1}=0.6,\ p_{2,2}^{a_1}=0.4,\ p_{1,1}^{a_2}=0.4,\ p_{1,2}^{a_2}=0.6,\ p_{2,1}^{a_2}=0.5,\ p_{2,2}^{a_2}=0.5;$$

$$c_{1,1}^{a_1}=0.7,\ c_{1,2}^{a_1}=0.3,\ c_{2,1}^{a_1}=0.6,\ c_{2,2}^{a_1}=0.4,\ c_{1,1}^{a_2}=0.4,\ c_{1,2}^{a_2}=0.6,\ c_{2,1}^{a_2}=0.5,\ c_{2,2}^{a_2}=0.5.$$

We consider the problem of finding the optimal stationary strategy for the corresponding Markov decision problem with minimal average cost and an arbitrary fixed starting state.

The data concerned with the actions in the considered Markov decision problem can be represented in a suitable form using the probability matrices

$$P^{a_1}=\begin{pmatrix}0.7 & 0.3\\ 0.6 & 0.4\end{pmatrix},\quad P^{a_2}=\begin{pmatrix}0.4 & 0.6\\ 0.5 & 0.5\end{pmatrix}$$

and the matrices of transition cost

$$C^{a_1}=\begin{pmatrix}1 & 0\\ -2 & 5\end{pmatrix},\quad C^{a_2}=\begin{pmatrix}0 & 4\\ 2 & -3\end{pmatrix}.$$

On fig. 1 this Markov process is represented by the multigraph $GM=(X,EM)$ with the set of vertices $X=\{1,2\}$. The set of directed edges EM contains parallel directed edges that correspond to probability transitions from one state to another for different actions.

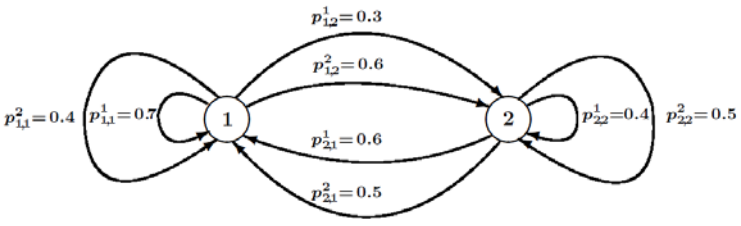

Fig.1

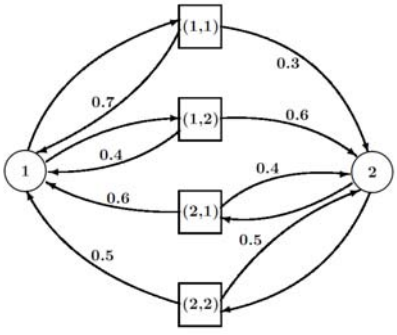

Fig.2

On fig. 2 is represented the graph $G'=(X',E')$. In G' the sets X_1', X_2', E_1', E_2' are defined as follows:

$$X_1'=\{1,2\},\ X_2'=X^1\cup X^2=\{(1,1),(1,2),(2,1),(2,2)\}$$

where

$$X^1=\{(1,1),(1,2)\},\ X^2=\{(2,1),(2,2)\}$$

and

$$E_1'=\{(1,(1,1)),(1,(1,2)),(1,(2,1)),(1,(2,2)),\ (3,(1,1)),(2,(1,2)),(2,(2,1)),(2,(2,2))\},$$

$$E_2'=\{((1,1),1),((1,1),2),((2,1),1),((2,2),1),\ ((1,2),2),((2,1),2),(2,2),2)\}.$$

The probabilities $p'_e = p'_{(x,a),y} = p^a_{x,y}$ for directed edges $((x,a),y) \in E'_2$ are written along the edges in fig. 2; and the costs of directed edges from E' are defined in the following way:

$$c_{1,(1,1)} = c_{1,(1,2)} = c_{1,(2,1)} = c_{1,(2,2)} = 0,\ c_{2,(1,1)} = c_{2,(1,2)} = c_{2,(2,1)} = c_{2,(2,2)} = 0,$$

$$c_{(1,1),1} = 1,\ c_{(1,1),2} = 0,\ c_{(2,1),1} = -2,\ c_{(2,2),1} = 2,\ c_{(1,2),1} = 0,\ c_{(1,2),2} = 4,\ c_{(2,1),2} = 5,\ c_{(2,2),2} = -3.$$

The set of possible stationary strategies for this Markov decision process consists of four strategies, i.e. $S = \{s^1, s^2, s^3, s^4\}$ where

$$s^1 : 1 \to a_1,\ 2 \to a_1;\ s^2 : 1 \to a_1,\ 2 \to a_2;\ s^3 : 1 \to a_2,\ 2 \to a_1;\ s^4 : 1 \to a_2,\ 2 \to a_2.$$

A fixed strategy s in Markov decision process generates a simple Markov process with transition costs, where the corresponding matrices P^s, C^s are formed from the rows of the matrices P^{a_i} and $C^{a_i}, i = 1,2$. As an example, if we fix the strategy s_2 then we obtain a simple Markov process with transition costs generated by the following matrices P^{s_2} and C^{s_2}:

$$P^{s_2} = \begin{pmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{pmatrix},\quad C^{s_2} = \begin{pmatrix} 1 & 0 \\ 2 & -3 \end{pmatrix}.$$

It easy to check that this Markov process is ergodic and the limit matrix of this process is

$$Q^{s_2} = \begin{pmatrix} 5/8 & 3/8 \\ 5/8 & 3/8 \end{pmatrix}.$$

The components of the vector of immediate costs $\mu^{s_2} = \begin{pmatrix} \mu_1^{s_2} \\ \mu_2^{s_2} \end{pmatrix}$ we can determine using formula $\mu_i^{s_2} = p_{i,1}^{s_2} c_{i,1}^{s_2} + p_{i,2}^{s_2} c_{i,2}^{s_2},\ i = 1,2,$ i.e. $\mu_1^{s_2} = 0.7$ and $\mu_2^{s_2} = 0.5$. In such a way we determine $f_1(s_2) = f_2(s_2) = 1/4$. Analogically can be calculated $f_1(s_1) = f_2(s_1) = 22/30$, $f_1(s_3) = f_2(s_3) = = 16/10$ and $f_1(s_4) = f_2(s_4) = 9/11$. We can see that the optimal stationary strategy for the Markov decision problem with minimial average cost criterion is s^2. This strategy can be found by solving the following linear programming problem on auxiliary network (G', X'_1, X'_2, p', c'):

Minimize

$$\bar{\psi}(\alpha, q) = 0.7q_{1,1} + 2.4q_{1,2} + 0.8q_{2,1} - 0.5q_{2,2}$$

subject to

$$\begin{cases} 0.7q_{1,1} + 0.4q_{1,2} + 0.6q_{2,1} + 0.5q_{2,2} = q_1, \\ 0.3q_{1,1} + 0.6q_{1,2} + 0.4q_{2,1} + 0.5q_{2,2} = q_2, \\ \alpha_{1,(1,1)} = q_{1,1}, \\ \alpha_{1,(1,2)} = q_{1,2}, \\ \alpha_{2,(2,1)} = q_{2,1}, \\ \alpha_{2,(2,2)} = q_{2,2}, \\ q_{1,1} + q_{1,2} + q_{2,1} + q_{2,2} + q_1 + q_2 = 1, \\ \alpha_{1,(1,1)}, \alpha_{1,(1,2)}, \alpha_{2,(2,1)}, \alpha_{2,(2,2)} \ge 0, \\ q_{1,1}, q_{1,2}, q_{2,1}, q_{2,2}, q_1, q_2 \ge 0. \end{cases}$$

The optimal solution of this problem is

$$q_1^* = 5/8,\quad q_2^* = 3/8,\quad q_{1,1}^* = 5/8,\quad q_{2,2}^* = 3/8,\quad q_{1,2}^* = 0,\quad q_{2,1}^* = 0,$$

$$\alpha_{1,(1,1)}^* = 5/8,\quad \alpha_{2,(2,2)}^* = 3/8,\quad \alpha_{1,(1,2)}^* = 0,\quad \alpha_{2,(2,1)}^* = 0.$$

The optimal value of the objective function is $\overline{\psi}(\alpha^*, q^*) = 1/4$.

We can find the optimal strategy on G':

$$s^*_{1,(1,1)} = 1, \quad s^*_{1,(1,2)} = 0, \quad s^*_{2,(2,1)} = 0, \quad s^*_{2,(2,2)} = 1.$$

This mean that the optimal stationary strategy for Markov decision problem is

$$s^* : 1 \to a_1, \quad 2 \to a_2$$

and the average cost per transition is $f_1(s^*) = f_2(s^*) = 1/4$.

The auxiliary graph with distinguished optimal strategies in the controllable states $x_1 = 1$ and $x_2 = 2$ is represented on fig. 3. The unique outgoing directed edge $(1,(1,1))$ from vertex 1 that end in vertex $(1,1)$ corresponds to the optimal strategy $1 \to a_1$ in the state $x = 1$ and the unique outgoing directed edge $2,(2,2)$ from vertex 2 that end in vertex $(2,2)$ corresponds to the optimal strategy $2 \to a_2$ in the state $x = 2$.

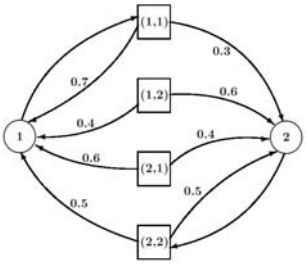

Fig.3

3. Linear Programming Approach for Average Markov Decision Problem and Algorithm for Determining the Optimal Strategies

In previous section we have shown that the optimal stationary strategies for Markov decision processes can be found by constructing an auxiliary stochastic control problem and applying the linear programming algorithm for the control problem on an auxiliary network. Below we show how to apply linear programming algorithm directly to Markov decision problem with average cost optimization criterion without construction the auxiliary stochastic control problem.

At first we describe the linear programming algorithm for a special class of Markov decision processes. We consider Markov decision processes with the property that an arbitrary stationary strategy $s : X \to A$ generate a recurrent Markov chain, i.e. we assume that the graph $GR^s = (X, GE^s)$ of the matrix of probability transition $P^s = (p_{x,y})^s$ is strongly connected. In general we can see that the linear programming approach can be used for an arbitrary Markov decision problems where an arbitrary stationary strategy generates a unichain. Such Markov decision processes we call *perfect Markov decision processes*. It is easy to observe that if for an arbitrary strategy $s : A \to X$ in the Markov decision process each row of the matrix $P^s = (p_{x,y})$ contains at least $\left[|X|/2\right]+1$ nonzero elements then Markov decision process is perfect, i.e. the corresponding graph $GR^s = (X, ER^s)$ is strongly connected.

Let $s : X \to A$ be an arbitrary strategy $(s \in S)$ for Markov decision process. Then for every fixed $x \in X$ we have a unique action $a = s(x) \in A(x)$ and therefore we can identify the map s with the set of boolean values $s_{x,a}$ for $x \in X$ and $a \in A(x)$, where

$$s_{x,a} = \begin{cases} 1, & \text{if} \quad a = s(x), \\ 0, & \text{if} \quad a \neq s(x). \end{cases}$$

In a similar way for the optimal stationary strategy s^* we shall with the boolean values $s^*_{x,a}$.

Assume that the Markov decision process is perfect. Then the following lemma holds.

Lemma 1. *A stationary strategy s^* is optimal if and only if it corresponds to an optimal solution of the following mixed integer bilinear programming problem:*

Minimize

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

subject to

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

where $\mu_{x,a} = \sum\limits_{y\in X} c^a_{x,y}\, p^a_{x,y}$ *is the immediate cost in the state* $x \in X$ *for a fixed action* $a \in A(x)$.

Proof. For a fixed strategy s the system (2) has a unique solution with respect q_x, $x \in X$ which represents the limiting probabilities of the recurrent Markov chains with the matrix of probability transition P^s. The value objective function (1) for this solution expresses the average cost per transition for an arbitrary fixed starting state. Therefore for fixed strategy s we have $f_x(s) = \psi(s, q^s), \ \forall x \in X$. This means that if we solve the optimization problem (1), (2) for the perfect Markov decision process then we obtain the optimal stationary strategy s^*.

Remark 1. For the perfect Markov decision processes the objective function $\psi(s,q)$ on the set of feasible solution depend only on $s_{x,a}$ for $x \in X, a \in A(x)$. Moreover the conditions $q_x \geq 0$ for $x \in X$ in (2) holds if $s_{x,a} \geq 0, \forall x \in X, a \in A(x)$ and therefore in the case of perfect Markov processes can be omitted. The conditions $q_x \geq 0, \forall x \in X$ in (2) are essential for non perfect Markov processes.

Based on Lemma 1 we can prove the following result.

Theorem 1. *Let* $\alpha^*_{x,y}$ $(x \in X_1, y \in X)$, q^*_x $(x \in X)$ *be a basic optimal solution of the following linear programming problem:*

Minimize

$$\bar{\psi}(\alpha, q) = \sum_{x\in X}\sum_{a\in A(x)} \mu_{x,a}\alpha_{x,a} \tag{3}$$

subject to

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

where $\mu_{x,a} = \sum\limits_{y\in X} c^a_{x,y}\, p^a_{x,y}$ *for* $x \in X$. *Then the optimal stationary strategy* s^* *for perfect Markov process can be found as follows:*

$$s^*_{x,a} = \begin{cases} 1, & \text{if} \quad \alpha^*_{x,a} > 0, \\ 0, & \text{if} \quad \alpha^*_{x,a} = 0, \end{cases}$$

where $x \in X, a \in A(x)$. *Moreover, for every starting state* $x \in X$ *the optimal average cost per transition is equal to* $\bar{\psi}(\alpha^*, q^*)$, *i.e.* $f_x(s^*) = \sum\limits_{x\in X}\sum\limits_{a\in A(x)} \mu_{x,a}\alpha^*_{x,a}$ *for every* $x \in X$.

Proof. We show that the bilinear programming problem (1), (2) with boolean variables $s_{x,a}$ for $x \in X_1, a \in A(x)$ can be reduced to the linear programming problem (3), (4). We observe that the restriction $s_{x,a} \in \{0,1\}$ in problems (1), (2) can be changed by $s_{x,a} \geq 0$ because the optimal basic solutions after such a transformation of the problem are not changed. In addition the restrictions

$$\sum_{a \in A(x)} s_{x,a} = 1, \ \forall x \in X,$$

can be changed by the restrictions $\sum_{a \in A(x)} s_{x,y} q_x = q_x, \ \forall x \in X$ because for the perfect Markov process holds $q_x > 0, \ \forall x \in X$. This means that the system (2) in the problem (1), (2) can be changed by the following system

$$\begin{cases} \sum_{x \in X} \sum_{a \in A(x)} p_{x,y}^a s_{x,a} q_x = q_y, \quad \forall y \in X, \\ \sum_{x \in X} q_x = 1, \\ \sum_{a \in A(x)} s_{x,a} q_x = q_x, \quad \forall x \in X, \\ s_{x,a} \geq 0, \ \forall x \in X, a \in A(x); \quad q_x \geq 0, \ \forall x \in X. \end{cases} \tag{5}$$

In a such way we may conclude that problem (1), (2) and problem (1), (5) have the same optimal solutions. Taking into account that for the perfect network $q_x > 0, \ \forall x \in X$ we can introduce in problem (1), (5) the notations $\alpha_{x,a} = s_{x,a} q_x$ for $x \in X, a \in A(x)$. In a such way we obtain the problem (3), (4). It is evident that $\alpha_{x,a} \neq 0$ if and only if $s_{x,y} = 1$. Therefore the optimal stationary strategy s^* can be found according to the rule formulated in the theorem.

So, if the Markov decision process is perfect then the optimal stationary strategy s^* can be found using the algorithm described below. It easy to observe that q_x in the system (4) can be eliminated if we take into account that $\sum_{a \in A(x)} \alpha_{x,a} = q_x, \quad \forall x \in X$. Then Theorem 1 we can formulate in the following way.

Theorem 2. *Let $\alpha_{x,y}^*$ $(x \in X, y \in X)$ be a basic optimal solution of the following linear programming problem:*

Minimize

$$\bar{\psi}(\alpha) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha_{x,a} \tag{6}$$

subject to

$$\begin{cases} \sum_{x \in X} \sum_{a \in A(x)} p_{x,y}^a \alpha_{x,a} - \sum_{a \in A(y)} \alpha_{y,a} = 0, \ \forall y \in X, \\ \sum_{x \in X} \sum_{a \in A(x)} \alpha_{x,a} = 1, \\ \alpha_{x,a} \geq 0, \ \forall x \in X, a \in A(x). \end{cases} \tag{7}$$

Then the optimal stationary strategy s^* for perfect Markov process can be found as follows:

$$s_{x,a}^* = \begin{cases} 1, & \text{if} \quad \alpha_{x,a}^* > 0, \\ 0, & \text{if} \quad \alpha_{x,a}^* = 0, \end{cases}$$

where $x \in X, a \in A(x)$. Moreover, for every starting state $x \in X$ the optimal average cost per transition is equal to $\bar{\psi}(\alpha^*, q^*)$, i.e. $f_x(s^*) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha_{x,a}^*$ for every $x \in X$.

Thus, the optimal stationary strategy for Markov decision problem can be found using the following algorithm [4, 5].

Algorithm. Determining the Optimal Stationary Strategies for Perfect Markov Decision Problem

1) Form the linear programming problem (3), (4) and find a basic optimal solution $\alpha^*_{x,y}, q^*_x$ of this problem;
2) Fix $s^*_{x,a} = 1$ for (x,a) that corresponds to the basic components of the optimal solution.

Example. Consider the Markov decision problem with average cost criterion from section 2. The corresponding multigraph of the Markov decision process is represented on fig. 1. The optimal stationary strategy s^* of this problem can be found by solving the linear programming problem (3), (4), i.e.:

Minimize

$$\bar{\psi}(\alpha, q) = 0.7\alpha_{1,1} + 2.4\alpha_{1,2} + 0.8\alpha_{2,1} - 0.5\alpha_{2,2}$$

subject to

$$\begin{cases} 0.7\alpha_{1,1} + 0.6\alpha_{2,1} + 0.4\alpha_{1,2} + 0.5\alpha_{2,2} = q_1, \\ 0.3\alpha_{1,1} + 0.4\alpha_{2,1} + 0.6\alpha_{1,2} + 0.5\alpha_{2,2} = q_2, \\ q_1 + q_2 = 1, \\ \alpha_{1,1} + \alpha_{1,2} = q_1, \\ \alpha_{2,1} + \alpha_{2,2} = q_2, \\ \alpha_{1,1}, \alpha_{1,2}, \alpha_{2,1}, \alpha_{2,2} \geq 0, \quad q_1, q_2 \geq 0. \end{cases}$$

The optimal solution of this problem is

$$q^*_1 = 5/8, \quad q^*_2 = 3/8, \quad \alpha^*_{1,1} = 5/8, \quad \alpha^*_{2,2} = 3/5, \quad \alpha^*_{1,2} = 0, \quad \alpha^*_{2,1} = 0$$

and the corresponding average cost is equal to $1/4$, i.e. $\bar{\psi}(\alpha^*, q^*) = 1/4$.

The optimal solution of the problem corresponds to the optimal stationary strategy $s^*_{1,1} = 1, \ s^*_{1,2} = 0, s^*_{2,1} = 0, \ s^*_{2,2} = 1$ i.e. $s^* : 1 \to a_1, \ 2 \to a_2$. So, optimal stationary strategy s^* determine the Markov process with the following probability and cost matrices $P^{s^*} = \begin{pmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{pmatrix}$, $C^{s^*} = \begin{pmatrix} 1 & 0 \\ 2 & -3 \end{pmatrix}$.

The graph of transition probabilities of this Markov process is represented on fig. 4.

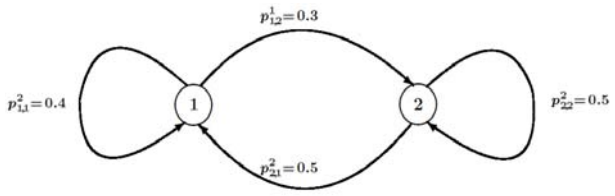

Fig.4

The result described above shows that the Markov decision problem with average cost criterion can be transformed into a stochastic optimal control problem on auxiliary network $(G', X_1, X_2, p', c', x_{i_0})$. This means that the linear programming algorithm can be developed and specified for Markov decision problems with average and discounted costs optimization criteria.

Bibliography:

1. Howard R.A., Dynamic Programming and Markov Processes. - Wiley, 1960.
2. Puterman M., Markov Decision Processes: Stochastic Dynamic Programming. - John Wiley, New Jersey, 2005.
3. White D.J. Markov Decision Processes. - Wiley, New York, 1993.
4. Lozovanu D, Pickl S., Optimal Stationary Control of Discrete Processes and a Polynomial Time Algorithm for Stochastic Control Problem on Networks, Proceedings of ICCS 2010 conference, Amsterdam, Elsevier Procedia Computer Science, 1-10, 2010.
5. Capcelea M., Linear programming approach for stochastic discrete optimal control problems with infinite time horizon, Proceedings of ISCOPAM conference, Iasi, July 12-16, 2010, p.7.

Prezentat la 11.11.2010