Determining Optimal Stationary Strategies for Discounted Stochastic Optimal Control Problem on Networks

Dmitrii Lozovanu[a,*], Stefan Pikl[b]

[a]*Institute of Mathematics and Computer Science, Academy of Sciences, Academy str., 5, Chisinau, MD–2028, Moldova*

[b]*Institut für Theoretische Informatik, Mathematik und Operations Research, Fakultät fur Informatik, Universität der Bundeswehr, München*

Abstract

The stochastic version of discrete optimal control problem with infinite time horizon and discounted integral-time cost criterion is considered. This problem is formulated and studied on certain networks. A polynomial time algorithm for determining the optimal stationary strategies for the considered problems is proposed and some applications of the algorithm for related Markov decision problems are described.

Key words: Discounted Stochastic Control Problem, Optimal Stationary Strategies, Polynomial Time Algorithm, Discounted Markov Processes

1 Introduction, Problem Formulation and the Main Concept

In this paper we consider the stochastic version of the following discrete optimal control problem with infinite time horizon and a discounted integral-time cost criterion by trajectory. Let a time-discrete system L with a finite set of states X be given. Assume that the dynamics of the system is described by a directed graph of states transitions $G = (X, E)$ where the set of vertices X corresponds to the set of states of the dynamical system; an arbitrary directed edge $e = (x, y)$ expresses the possibility of the system to pass from the state $x = x(t)$ to the state $y = x(t)$ at every discrete moment of time

* Dmitrii Lozovanu
Email address: `lozovanu@usm.md, stefan.pickl@unibw.de` (Stefan Pikl).

$t = 0, 1, 2, \ldots$. Hereby, a directed edge $e = (x, y) \in E$ corresponds to a feasible stationary control of system L in the state x and the subset of edges $E^+(x) = \{e = (x, y) \in E | y \in X\}$ corresponds to the set of feasible stationary controls of the system in the state $x \in X$. We assume that on the edge set E a cost function $c : E \to R$ is defined which assigns a cost c_e to each directed edge $e = (x, y) \in E$ when the system makes a transition from the state $x = x(t)$ to the state $y = x(t+1)$ for every $t = 0, 1, 2, \ldots$,i.e. the costs $c_{x(t),x(t+1)}$ does not depend on t. We define a stationary control of system L in G as a map

$$s : x \to y \in X^+(x) \quad for \quad x \in X,$$

where $X^+(x) = \{y \in X | (x, y) \in E\}$. Let s be an arbitrary stationary control. Then the set of edges of the form $(x, s(x))$ in G generates a subgraph $G_s = (X, E_s)$ where each vertex $x \in X$ contains one leaving directed edge. So, if the starting state $x_0 = x(0)$ is fixed then the system makes transitions from one state to another through the corresponding directed edges $e_0^s, e_1^s, e_2^s, \ldots, e_t^s, \ldots$, where $e_t^s = (x(t), x(t+1)), \quad t = 0, 1, 2, \ldots$. This sequence of directed edges generates a trajectory $x_0 = x(0), x(1), x(2), \ldots$ which leads to a unique directed cycle. For an arbitrary stationary strategy s and a fixed starting state x_0 the discounted integral-time cost $\sigma_{x_0}^{\lambda}(s)$ is defined as follows $\sigma_{x_0}^{\lambda}(s) = \sum_{t=0}^{\infty} \lambda^t c_{e_t^s}$, where $\lambda, \quad 0 \le \lambda < 1$, is a given (so called) discounted factor. Based on the results from [1,3] it is easy to show that for an arbitrary stationary strategy s there exists $\sigma_{x_0}^{\lambda}(s)$. If we denote by $\sigma^{\lambda}(s)$ the vector column with components $\sigma_x^{\lambda}(s)$ for $x \in X$ then $\sigma_{x_0}^{\lambda}(s)$ can be found by solving the system of linear equations $(I - \lambda P^s)\sigma^{\lambda}(s) = \overline{c}^s$, where $\overline{c}^s$ is the vector with corresponding components $c_{(x,s(x))}$ for $x \in X, \quad I$ is the identity matrix and P^s the matrix with elements $p_{x,y}^s$ for $x, y \in X$ defined as follows

$$p_{x,y}^s = \begin{cases} 1, \text{ if } y = s(x); \\ 0, \text{ if } y \neq s(x). \end{cases}$$

We are seeking for a stationary control s^* such that $\sigma_{x_0}^{\lambda}(s^*) = \min_s \sigma_{x_0}^{\lambda}(s)$. In this paper we consider the stochastic version of the problem formulated above. We assume that the dynamical system may admit states in which the vector of control parameters is changed in a random way. So, the set of states X is divided into two subsets $X = X_1 \cup X_2, \quad X_1 \cap X_2 = \emptyset$, where X_1 represents the set of states in which the decision maker is able to control the dynamical system and X_2 represents the set of states in which the dynamical system makes transition to the next state in a random way. So, for every $x \in X$ on the set of feasible transitions $E^+(x)$ the distribution function $p : E^+(x) \to R$ is defined such that $\sum_{e \in E^+(x)} p_e = 1, \quad p_e \ge 0, \forall e \in E^+(x)$ and the transitions from the states $x \in X_2$ to the the next states are made according to these distribution functions. Here in a similar way as in the previous case of the problem we assume that to each directed edge $e = (x, y) \in E$ a cost c_e is associated when the system makes a transition from the state $x = x(t)$ to the state $y = x(t+1)$ for every $t = 0, 1, 2, \ldots$. In addition we assume that the discounted factor λ, $0 \le \lambda < 1$, and the starting state x_0 are given. We define a stationary control for the considered problem as a map

$$s : x \to y \in X^+(x) \quad for \quad x \in X_1.$$

For an arbitrary stationary strategy s we define the graph $G_s = (X, E_s \cup E_{X_2})$, where $E_s = \{e = (x, y) \in E | x \in X_1, y = s(x)\}$, $E_{X_2} = \{e = (x, y) | x \in X_2, y \in X\}$. This graph corresponds to a Markov process with the probability matrix $P^s = (p^s_{x,y})$, where

$$p^s_{x,y} = \begin{cases} p_{x,y}, \text{ if } x \in X_2 \text{ and } y = X; \\ 1, \quad \text{ if } x \in X_1 \text{ and } y = s(x); \\ 0, \quad \text{ if } x \in X_1 \text{ and } y \neq s(x). \end{cases}$$

For this Markov process with associated costs c_e, $e \in E$ we can define the expected discounted integral-time cost $\sigma^\lambda_{x_0}(s)$ in the same way as for discounted Markov processes with rewards (if we treat the rewards as the costs). In this paper we consider the problem of determining the strategy s^* for which $\sigma^\lambda_{x_0}(s^*) = \min_s \sigma^\lambda_{x_0}(s)$.

2 The Main Results

The stationary case of the considered discounted stochastic control problem can be studied and solved using the general concept of Markov decision processes and the linear programming approach to corresponding problems (see [1–3]). Here we develop a new technique and we will formulate a new linear programming problem which is more suitable to the specific context. To obtain our linear model we shall use the following condition:

$$\begin{cases} \sigma_x - \lambda \sum_{y \in X^+(x)} p^s_{x,y} \sigma_y = \sum_{y \in X^+(x)} c_{(x,y)} p^s_{x,y}, \quad \forall x \in X_1; \\ \sigma_x - \lambda \sum_{y \in X^+(x)} p_{x,y} \sigma_y = \sum_{y \in X+(x)} c_{(x,y)} p_{x,y}, \quad \forall x \in X_2, \end{cases} \tag{1}$$

for an arbitrary stationary strategy s. For fixed s the probabilities $p^s_{x,y}$, $x \in X, y \in X^+(x)$, satisfy the conditions: $\sum_{y \in X^+(x)} p^s_{x,y} = 1, \forall x \in X_1$; $p_{x,y} \in \{0, 1\}, \forall x \in X_1, y \in X^+(x)$. The system (1) has a unique solution with respect to σ_x for $x \in X$ and therefore we uniquely determine $\sigma^s_{x_0}$. Thus we can consider the linear programming problem: Maximize

$$\psi_{p^s}(\sigma) = \sigma_{x_0} \tag{2}$$

subject to (1). This problem has a unique feasible solution which is the optimal one. The dual program for this problem is: Minimize

$$\varphi_{p^s}(\alpha) = \sum_{x \in X_1} \sum_{y \in X(x)} c_{(x,y)} p^s_{x,y} \alpha_x + \sum_{x \in X_2} \sum_{y \in X(x)} c_{(x,y)} p_{x,y} \alpha_x \tag{3}$$

subject to

$$\begin{cases} \alpha_y - \lambda \sum_{x \in X_1^-(y)} p^s_{x,y} \alpha_x - \lambda \sum_{x \in X_2^-(y)} p_{x,y} \alpha_x \geq 1, \quad y = x_0; \\ \alpha_y - \lambda \sum_{x \in X_1^-(y)} p^s_{x,y} \alpha_x - \lambda \sum_{x \in X_2^-(y)} p_{x,y} \alpha_x \geq 0, \quad \forall\, y \in X \setminus \{x_0\}. \end{cases} \tag{4}$$

If we take here the minimum with respect to s then we obtain a bilinear programming problem with respect to α_x and $p^s_{x,y}$, where $p^s_{x,y}$ satisfy the conditions: $\sum_{y \in X^+(x)} p^s_{x,y} = 1; p^s_{x,y} \in \{0,1\}, \forall x \in X_1, y \in X^+(x)$. We have proved that the optimal solution is preserved if these conditions are changed by conditions: $\sum_{y \in X^+(x)} p^s_{x,y} \alpha_x = \alpha_x, \forall x \in X_1, \forall y \in X^+(x); \alpha_x \geq 0, \beta_{x,y} \geq 0, \forall x \in X, y \in X^+(y)$. If we substitute after that operation $\beta_{x,y} = p^s_{x,y} \alpha_x$ then our bilinear programming problem obtained on the bases of (3),(4) with mentioned above conditions is reduced to the linear programming problem: Minimize

$$\varphi(\alpha, \beta) = \sum_{x \in X_1} \sum_{y \in X(x)} c_{(x,y)} \beta_{x,y} + \sum_{x \in X_2} \sum_{y \in X(x)} c_{(x,y)} p_{x,y} \alpha_x \tag{5}$$

subject to

$$\begin{cases} \alpha_y - \lambda \sum_{x \in X_1^-(y)} \beta_{x,y} - \lambda \sum_{x \in X_2^-(y)} p_{x,y} \alpha_x \geq 1, \quad y = x_o; \\ \alpha_y - \lambda \sum_{x \in X_1^-(y)} \beta_{x,y} - \lambda \sum_{x \in X_2^-(y)} p_{x,y} \alpha_x \geq 0, \quad \forall\, y \in X \setminus \{x_0\}; \\ \sum_{y \in X^-(x)} \beta_{x,y} = \alpha_x, \quad \forall x \in X_1;\ \beta_{x,y} \geq 0, \alpha_x \geq 0, \forall x \in X, y \in X^+(x), \end{cases} \tag{6}$$

where $X_1^-(y) = \{x \in X_1 | (x,y) \in E\}$, $X_2^-(y) = \{x \in X_2 | (x,y) \in E\}$. The following result holds: If $\alpha^*_x, \beta^*_{x,y}$ is a basic optimal solution of the problem (5),(6) and $\alpha^*_x \neq 0$ for $x \in X_1$ then $p^{s^*}_{x,y} = \beta^*_{x,y}/\alpha^*_x \in \{0,1\}, x \in X_1, y \in X^+(y)$; the optimal stationary strategy $s^* : x \to y$ for $y \in X^+(x)$ corresponds to $p^{s^*}_{x,y} = 1$ for $x \in X_1, y \in X^+(x)$.

References

[1] Howard R.A., Dynamic Programming and Markov Processes. Wiley (1960)

[2] Lozovanu D., Pickl. S., Optimization and Multiobjective Control of Time-Discrete Systems. Springer Verlag (2009)

[3] Puterman M., Markov Decision Processes. Wiley (1993)