BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 2(60), 2009, Pages 73–90
ISSN 1024–7696

Dynamic Programming Algorithms for Solving Stochastic Discrete Control Problems

Dmitrii Lozovanu, Stefan Pickl

Abstract. The stochastic versions of classical discrete optimal control problems are formulated and studied. Approaches for solving the stochastic versions of optimal control problems based on concept of Markov processes and dynamic programming are suggested. Algorithms for solving the problems on stochastic networks using such approaches and time-expended network method are proposed.

Mathematics subject classification: 93E20, 49L20.
Keywords and phrases: Time-Discrete System, Optimal Control, Stochastic Networks, Markov Processes, Dynamic Programming.

1 Introduction

The paper is concerned with studying and solving the stochastic versions of the classical discrete optimal control problems from [1,2,5]. In the deterministic control problems from [1,2] the choosing of the vector of control parameters from the corresponding feasible set at every moment of time for an arbitrary state is assumed to be at our disposition, i.e each dynamical state of the system is assumed to be controllable. In this paper we consider the control problems for which the discrete system in the control process may meet dynamical states where the vector of control parameters is changing in a random way according to given distribution functions of the probabilities on given feasible dynamical stats. We call such states of dynamical system uncontrollable dynamical states. So, we consider the control problems for which the dynamics may contain controllable states as well uncontrollable ones. We show that in general form these versions of the problems can be formulated on stochastic networks and new approaches for their solving based on concept of Markov processes and dynamic programming from [3,4] can be suggested. Algorithms for solving the considered stochastic versions of the problems using the mentioned concept and the time-expended network method from [5,6] are proposed and grounded.

2 Problems Formulations and the Main Concept

We consider a time-discrete system L with a finite set of states $X \subset R^n$. At every time-step $t = 0, 1, 2, \dots,$ the state of the system L is $x(t) \in X$. Two states x_0 and x_f are given in X, where $x_0 = x(0)$ represents the starting state of system

L and x_f is the state in which the system L must be brought, i.e. x_f is the final state of L. We assume that the system L should reach the final state x_f at the time-moment $T(x_f)$ such that $T_1 \le T(x_f) \le T_2$, where T_1 and T_2 are given. The dynamics of the system L is described as follows

$$x(t+1) = g_t(x(t), u(t)), \quad t = 0, 1, 2, \dots, \tag{1}$$

where

$$x(0) = x_0 \tag{2}$$

and $u(t) = (u_1(t), u_2(t), \dots, u_m(t)) \in R^m$ represents the vector of control parameters. For any time-step t and an arbitrary state $x(t) \in X$ a feasible finite set $U_t(x(t)) = \{u^1_{x(t)}, u^2_{x(t)}, \dots, u^{k(x(t))}_{x(t)}\}$, for the vector of control parameters $u(t)$ is given, i.e.

$$u(t) \in U_t(x(t)), \quad t = 0, 1, 2, \dots . \tag{3}$$

We assume that in (1) the vector functions $g_t(x(t), u(t))$ are determined uniquely by $x(t)$ and $u(t)$, i.e. the state $x(t+1)$ is determined uniquely by $x(t)$ and $u(t)$ at every time-step $t = 0, 1, 2, \dots$. In addition we assume that at each moment of time t the cost $c_t\big(x(t), x(t+1)\big) = c_t\big(x(t), g_t(x(t), u(t))\big)$ of system's transaction from the state $x(t)$ to the state $x(t+1)$ is known.

Let $x_0 = x(0), x(1), x(2), \dots, x(t), \dots$ be a trajectory generated by given vectors of control parameters $u(0), u(1), \dots, u(t-1), \dots$. Then either this trajectory passes through the state x_f at the time-moment $T(x_f)$ or it does not pass through x_f.We denote

$$F_{x_0 x_f}(u(t)) = \sum_{t=0}^{T(x_f)-1} c_t\big(x(t), g_t(x(t), u(t))\big) \tag{4}$$

the integral-time cost of system's transactions from x_0 to x_f if $T_1 \le T(x_f) \le T_2$; otherwise we put $F_{x_0 x_f}(u(t)) = \infty$. In [1, 2, 5] have been formulated and studied the following problem: to determine the vectors of control parameters $u(0), u(1), \dots, u(t), \dots$ which satisfy conditions (1)-(3) and minimize functional (4). This problem can be regarded as a control model with controllable states of dynamical system because for an arbitrary state $x(t)$ at every moment of time the choosing of vector of control parameter $u(t) \in U_t(x(t))$ is assumed to be at our disposition. In the following we consider the stochastic versions of the control model formulated above. We assume that the dynamical system L may contain uncontrollable states, i.e. for the system L there exist dynamical states in which we are not able to control the dynamics of the system and the vector of control parameters $u(t) \in U_t(x(t))$ for such states is changing in the random way according to given distribution function

$$p : U_t(x(t)) \to [0, 1], \qquad \sum_{i=1}^{k(x(t))} p(u^i_{x(t)}) = 1 \tag{5}$$

on the corresponding dynamical feasible sets $U_t(x(t))$. If we regard arbitrary dynamic state $x(t)$ of system L at given moment of time t as position (x, t) then the set of positions

$$Z = \{(x,t) \,|\, x \in X,\ t = 0,1,2,\dots,T_2\}$$

of dynamical system can be divided into two disjoint subsets

$$Z = Z^C \cup Z^N \quad (Z^C \cap Z^N = \emptyset),$$

where Z^C represents the set of controllable positions of L and Z^N represents the set of positions $(x,t) = x(t)$ for which the distribution function (5) of the vectors of control parameters $u(t) \in U_t(x(t))$ are given. This mean that the dynamical system L works as follows. If the starting point belongs to controllable positions then the decision maker fixes a vector of control parameter and we obtain the state $x(1)$. If the starting state belongs to the set of uncontrollable positions then the system passes to the next state in a random way. After that if at the time-moment $t = 1$ the state $x(1)$ belong to the set of controllable positions then the decision maker fixes the vector of control parameter $u(t) \in U_t(x(t))$ and we obtain the state $x(2)$. If $x(1)$ belongs to the set of uncontrollable positions then the system passes to the next state in a random way and so on. In this dynamic process the final state may be reached at given moment of time with a probability which depend on the control of the system in the deterministic states as well as the expectation of integral time cost by trajectory depends on control of the system in these states. The main results of this paper are concerned with studying and solving the following problems.

1. For given vectors of control parameters $u(t) \in U_t(x(t))$, $x(t) \in Z^C$, to determine the probability that the dynamical system L with given starting state $x_0 = x(0)$ will reach the final state x_f at the moment of time $T(x_2)$ such that $T_1 \le T(x_f) \le T_2$. We denote this probability $P_{x_0}\big(u(t), x_f, T_1 \le T(x_f) \le T_2\big)$; if $T_1 = T_2 = T$ then we use the the notation $P_{x_0}(u(t), x_f, T)$.

2. To find the vectors of control parameters $u^*(t) \in U_t(x(t))$, $x(t) \in Z^C$ for which the probability in problem 1 is maximal. We denote this probability we denote $P_{x_0}\big(u^*(t), x_f, T_1 \le T(x_f) \le T_2\big)$; in the case $T_1 = T_2 = T$ we shall use the notation $P_{x_0}(u^*(t), x_f, T)$.

3. For given vectors of control parameters $u(t) \in U_t(x(t))$, $x(t) \in Z^C$ and given number of stages T to determine the expectation of integral-time of system's transactions within T stages for system L with staring state $x_0 = x(0)$. We denote this expectation $C_{x_0}(u(t), T)$.

4. To determine the vectors of control parameters $u^*(t) \in U_t(x(t))$, $x(t) \in Z^C$ for which the expectation of integral-time cost for dynamical system in problem 3 is minimal. We denote this expectation $C_{x_0}(u^*(t), x_f, T)$.

5. For given vectors of control parameters $u(t) \in U_t(x(t))$, $x(t) \in Z^C$, to determine the expectation of integral-time cost of system's transactions from starting state x_0 to final state x_f when the final state is reached at the time-moment $T(x_f)$ such that $T_1 \le T(x_f) \le T_2$. We denote this expectation $C_{x_0}\big(u(t), x_f, T_1 \le T(x_f) \le T_2\big)$; if $T_1 = T_2 = T$ then we denote $C_{x_0}(u(t), x_f, T)$.

6. To determine the vectors of control parameters $u^*(t) \in U_t(x(t))$, $x(t) \in Z^C$ for which the expectation of integral-time cost of system's transactions in problem 5 is minimal. We denote this expectation $C_{x_0}\big(u^*(t), x_f, T_1 \le T(x_f) \le T_2\big)$; in the case $T_1 = T_2 = T$ we shall use the notation $C_{x_0}(u^*(t), x_f, T)$.

It is easy to observe that problems 1-6 extend and generalize a large class of deterministic and stochastic dynamic problems including problems from [1,2,4]. The problems from [4] related to finite Markov processes became problems 1–3 in the case when $Z^C = \emptyset$ and the probabilities $p(u^i_{x(t)})$ do not depend on time but depend only on states; the discrete optimal control problems from [1, 2] became problems 4-6 in the case $Z^N = \emptyset$. In the following we propose algorithms for solving the problem formulated above based on results from [1, 2, 4] and time-expended method from [5, 6].

3 Some Auxiliary Results and Definitions of the Basic Notions

In this section we describe some auxiliary results concerned with calculation of the state probabilities in a simple finite Markov processes and make more precise some basic definitions for our control problems. We shall use these results and the specification of the basic notion we shall use in next sections for a strict argumentation of the algorithms for solving problems 1-6.

3.1 Determining the State Probabilities of the Dynamical System in Finite Markov Processes

We consider a dynamical system with the set of states X where for every state $x \in X$ are given the probabilities $p_{x,y}$ of system's passage from x to another states $y \in X$ such that $\sum_{y \in X} p_{x,y} = 1$. Here the probabilities $p_{x,y}$ do not depend on time, i.e. we have a simple Markov process determined by the stochastic matrix of probabilities $P = (p_{x,y})$ and the starting state x_0 of dynamical system. The probability $P_{x_0}(x, t)$ of system's passage from the state x_0 to an arbitrary state $x \in X$ by using given t unite of time is defined and calculated on the basis of the following recursive formula [4]

$$P_{x_0}(x, \tau + 1) = \sum_{y \in X} P_{x_0}(y, \tau) p_{y,x}, \quad \tau = 0, 1, 2, \dots, t$$

where $P_{x_0}(x_0, 0) = 1$ and $P_{x_0}(x, 0) = 0$ for $x \in X \setminus \{x_0\}$. In the case when the probabilities of system's passage from one state to another depend on time we have a non-stationary process defined by a dynamic matrix $P(t) = (p_{x,y}(t))$ which describe this process. If this matrix is stochastic for every moment of time $t = 0, 1, 2, \dots$, then the state probabilities $P_{x_0}(x, t)$ can be defined and calculated by using a formula obtained similarly from written one changing $p_{x,y}$ by $p_{x,y}(\tau)$.

Now let us show how to calculate the probability of systems passage from the state x_0 to the state x when x is reached at the time moment $T(x)$ such that $T_1 \le T(x) \le T_2$ where T_1 and T_2 are given. So, we are seeking for the probability that the system L will reach the state x at least at one of the moments of time

$T_1, T_1+1, \ldots, T_2$. We denote this probability $P_{x_0}(x, T_1 \leq T(x) \leq T_2)$. For this reason we shall give the graphical interpretation of the simple Markov processes by using the random graph of state transitions $GR = (X, ER)$. In this graph each vertex $x \in X$ corresponds to a state of dynamical system and a possible system passage from one state x to another state y with positive probability $p_{x,y}$ is represented by the directed edge $e = (x, y) \in ER$ from x to y; to directed edges $(x, y) \in ER$ in G the corresponding probabilities $p_{x,y}$ are associated. It is evident that in the graph GR each vertex x contains at least one leaving edge (x, y) and $\sum\limits_{y \in X} p_{x,y} = 1$. In general we will consider also the stochastic process which may stop if one of the states from given subset of states of dynamical system is reached. This means that the random graph of such process may contain the deadlock vertices. So, we consider the stochastic process for which the random graph may contains the deadlock vertices and $\sum\limits_{y \in X} p_{x,y} = 1$ for the vertices $x \in X$ which contain at least one leaving directed edge. Such random graphs do not correspond to Markov processes and the matrix of probability P contains rows with zero components. Nevertheless the probabilities $P_{x_0}(x, t)$ in the both cases of the considered processes can be calculated on the basis of recursive formula given above. In the next sections we can see that the state probabilities of the system can be also calculated starting from final state by using the backward dynamic procedure.In the following the random graph with given probability function $p : ER \to R$ on edge set ER and given distinguished vertices which correspond to starting and final states of dynamical system we be called the stochastic network. Further we shall use the stochastic networks for calculation of the probabilities $P_{x_0}(x, T_1 \leq T(x) \leq T_2)$.

Lemma 1. *Let be given a Markov process determined by stochastic matrix of probabilities $P = (p_{x,y})$ and the starting state x_0. Then the following formula holds:*

$$P_{x_0}(x, T_1 \leq T(x) \leq T_2) = P_{x_0}(x, T_1) + P_{x_0}^{T_1}(x, T_1+1) + \\ + P_{x_0}^{T_1, T_1+1}(x, T_1+1) + \cdots + P_{x_0}^{T_1, T_1+1, \ldots, T_2-1}(x, T_2) \qquad (6)$$

where $P_{x_0}^{T_1, T_1+1, \ldots, T_1+i-1}(x, T_1+i), i = 1, 2, \ldots, T_2 - T_1$, is the probability that the system L will reach the state x from x_0 by using $T_1 + i$ transactions and it does not pass through x at the moments of times $T_1, T_1+1, T_1+2, \ldots, T_1+i-1$.

Proof. Taking into account that $P_{x_0}(x, T_1 \leq T(x) \leq T_1 + i$ expresses the probability of the system L to reach from x_0 the state x at least at one of the moments of time $T_1, T_1+1, \ldots, T_1+i$ we can write the following recursive formula

$$P_{x_0}(x, T_1 \leq T(x) \leq T_1+i) = P_{x_0}(x, T_1 \leq T(x) \leq T_1+i-1) + \\ + P_{x_0}^{T_1, T_1+1, \ldots, T_1+i-1}(x, T_1+i). \qquad (7)$$

Applying $T_2 - T_1$ times this formula for $i = 1, 2, \ldots, T_2 - T_1$ we obtain the equality (6). □

Note that formula 6 and 7 couldn't be used directly for calculation of the probability $P_{x_0}(x, T_1 \le T(x) \le T_2)$. Nevertheless we can see that such representation of the probability $P_{x_0}(x, T_1 \le T(x) \le T_2)$ in the time expended network method will allow to ground a suitable algorithms for calculation of this probability and to develop new algorithms for solving problems from Section 2.

Corollary 1. *If the state x of dynamical system L in random graph $GR = (X, ER)$ corresponds to a deadlock vertex then*

$$P_{x_0}\big(x, T_1 \le T(x) \le T_2\big) = \sum_{t=T_1}^{T_2} P_{x_0}(x, t). \tag{8}$$

Let X_f be a subset of X and assume that at the moment of time $t = 0$ the system L is in the state x_0. Denote by $P_{x_0}\big(X_f, T_1 \le T(X_f) \le T_2\big)$ the probability that at least one of the states $x \in X_f$ will be reached at the time moment $T(x)$ such that $T_1 \le T(x) \le T_2$. Then the following corollary holds.

Corollary 2. *If the subset of states $X_f \subset X$ of dynamical system L in the random graph $GR = (X, ER)$ corresponds to the subset of deadlock vertices then for the probability $P_{x_0}\big(X_f, T_1 \le T(X_f) \le T_2\big)$ the following formula holds*

$$P_{x_0}\big(X_f, T_1 \le T(X_f) \le T_2\big) = \sum_{x \in X_f} \sum_{t=T_1}^{T_2} P_{x_0}(x, t). \tag{9}$$

3.2 Determining the Expectation of Integral-time cost of system's transactions in Finite Markov Processes

In order to define strictly the expectation of integral-time cost for dynamical system in problems 3-6 we need to introduce the notion of expectation of integral-time cost for finite Matkov processes with cost function on the set of state's transaction of dynamical system. We introduce this notion we introduce in the same way as the total expected earning in the Markov processes with rewards introduced in [4]. We consider a simple Marcov process determined by the stochastic matrix $p = (p_{x,y})$ and starting state x_0 of system L. Assume that for arbitrary two states $x, y \in X$ of the dynamical system is given the value $c_{x,y}$ which we treat as the cost of system L to pass from the state x to the state y. The matrix $C = (c_{x,y})$ is called the matrix of the costs of system's transactions for the dynamical system. Note that in [4] the values $c_{x,y}$ for given x are treated as the "earning" of the system's transaction from the state x to the states $y \in X$ and the corresponding Markov process with associated matrix C is called Markov process with reward. The Markov process with associated cost matrix C generates a sequence of costs when the system makes transactions from one state to another. Thus the cost is a random variable with a probability distribution induced by the probability relations of the Markov process. This means that for the system L the integral-time cost during T transactions is a random variable for which the expectation can be defined. We denote the expectation of integral-time cost in such process by $C_{x_0}(T)$. So, $C_{x_0}(T)$ expresses the expected integral-time cost of the system in the next T transactions if the system at the starting moment of time is in the state $x_0 = x(0)$. For an arbitrary $x \in X$ the values $C_x(\tau)$ are defined strictly and calculated on the basis of the following recursive formula

$$C_x(\tau) = \sum_{y \in Y} p_{x,y}(c_{x,y} + C_y(\tau - 1)), \quad \tau = 1, 2, \ldots, t$$

where $C_x(0) = 0$ for every $x \in X$. This formula can be treated in the similar way as formula for calculation the total earning in the Markov processes with rewards [4]. The expression $c_{x,y} + C_y(\tau - 1)$ means that if the system makes transaction from the state x to the state y then it spends the amount $c_{x,y}$ plus the amount it expects to spend during the next $\tau - 1$ transactions when the system start transactions in the state y at the moment of time $\tau = 1$. Taking into account that in the state x the system makes transactions in the random way with the probability distribution $p_{x,y}$ we obtain that the values $c_{x,y} + C_y(\tau - 1)$ should be weighted by the probabilites of transactions $p_{x,y}$. In the case of the non-stationary process, i.e. when the probabilities and the costs are changing in time, the expectation of integral-time cost of dynamical system is defined and calculated in similar way; in formula written above we should change $p_{x,y}$ by $p_{x,y}(\tau)$ and $c_{x,y}$ by $c_{x,y}(t)$.

3.3 Definition of the State Probability and The Expectation of Integral time cost in Control Problems 1–6

Using the definitions from previous subsections we can now specify the notions of state probabilities $P_{x(0)}(u(t), x, T)$, $P_{x(0)}\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$ and the expectations of integral-time cost $C_{x(0)}(u(t), T)$, $C_{x(0)}\big(u(t), x_f, T\big)$, $C_{x(0)}\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$ in problems 1-6. First of all we stress our attention to the definition of probability $P_{x_0}(u(t), x, T)$. For given starting state x_0, given time-moment T and fixed control $u(t)$ we define this probability in the following way. We consider that each system passage from an controllable state $x = x(t)$ to the next state $y = x(t+1)$ generated by the control $u(t)$ is made with probability $p_{x,y} = 1$ and the rest of probabilities of system's passages from x at the moment of time t to the next states are equal to zero. Thus we obtain a finite Markov process for which the probability of system passage from starting state x_o to final state x by using T unites of time can be defined. We denote this probability $P_{x_0}(u(t), x, T)$. We define the probability $P_{x_0}\big(u(t), x, T_1 \leq T(x) \leq T_2\big)$ for given T_1 and T_2 as probability of the dynamical system L to reach the state x at least at one of the moments of time $T_1, T_1 + 1, \ldots, T_2$. In order to define strictly the expectation of integral-time cost of dynamical system in problems 3-6 we shall use the notion of expectation of integral-time cost for Markov processes with costs defined on system's transactions. The expectation of integral-time cost $C_{x_0}(u(t), T)$ of system L in problem 3 for fixed control $u(t)$ is defined as the expectation of the integral-time cost during

T transitions of dynamical system in the Markov process generated by the control $u(t)$ and the corresponding costs of state's transactions of dynamical system. The expectation $C_{x_0}\big(u(t), x, T_1 \le T(x) \le T_2\big)$ in the problems 5 and 6 will made more precise in more detail form in Section 5.

4 The Main Approach and Algorithms for Determining the State Probabilities in the Control Problems on Stochastic Networks

In order to provide a better understanding of the main approach and to ground the algorithms for solving the problems formulated in Section 2 we shall use the network representation of the dynamics of the system and will formulate these problems on stochastic network. Note that in our control problems the probabilities and the costs of system's passage from one state to another depend on time. Therefore here we develop time-expended network method from [5, 6] for the stochastic versions of control problems and reduce them to the static cases of the problems. At first we show how to construct the stochastic network and how to solve the problems with fixed number of stages, i.e. we consider the case $T_1 = T_2 = T$.

4.1 Construction of Stochastic Network and Algorithms for Solving the Problems in the Case $T_1 = T_2 = T$

If the dynamics of discrete system L and the information related to the feasible sets $U_t(x(t))$ and the cost functions $c_t(x(t), g_t(x(t), u(t)))$ in the problems with $T_1 = T_2 = T$ are known then our stochastic network can be obtained in the following way. We identify each position (x, t) which correspond to a dynamic state $x(t)$ with a vertex $z = (x, t)$ of the network. So, the set of vertices Z of the network can be represented as follows $Z = Z_1 \cup Z_2 \cup \dots \cup Z_T$ where $Z_t = \{(x, t)\,|\, x \in X\}, t = 0, 1, 2, \dots, T$. To each vector of control parameters $u(t) \in U_t(x(t)), t = 1, 2, \dots, T-1$ which provide a system passage from the state $x(t) = (x, t)$ to the state $x(t+1) = (y, t+1)$ we associate in our network a directed edge $e(z, w) = \big((x, t), (y, t+1)\big)$ from the vertex $z = (x, t)$ to the vertex $w = (y, t+1)$, i.e., the set of edges E of the network is determined by the feasible sets $U_t(x(t))$. After that to each directed edge $e = (z, w) = \big((x, t), (y, t+1)\big)$ originating in uncontrollable positions (x, t) we put in correspondence the probability $p(e) = p\,(u(t))$, where $u(t)$ is the vector of control parameter which provide the passage of the system from the state $x = x(t)$ to the state $x(t+1) = (y, t+1)$. Thus if we distinguish in E the subset of edges $E_N = \big\{e = (z, w) \in E | z \in Z^N\big\}$ originating in uncontrollable positions Z^N then on E_N we obtain the probability function $p : E \to R$ which satisfies the condition

$$\sum_{e \in E^+(z)} p(e) = 1, \quad z \in Z^N \setminus Z_T$$

where $E^+(z)$ is the set of edges originating in z. In addition in the network we add to the edges $e = (z, w) = \big((x, t), (y, t+1)\big)$ the costs $c(z, w) = c\big((x, t), (y, t+1)\big) = c_t\big(x(t), x(t+1)\big)$ which correspond to the costs of system's passage from states $x(t)$ to the states $x(t+1)$. The subset of edges of the graph G originating in vertices

$z \in Z^C$ is denoted E_C, i.e. $E_C = E \setminus E_N$. So, our network is determined by the tuple $(G, Z^C, Z^N, z_0, z_f, c, p, T)$, where $G = (Z, E)$ is the graph which describes the dynamics of the system; the vertices $z_0 = (x, 0)$ and $z_f = (x_f, 0)$ correspond to the starting and the final states of the dynamical system, respectively; c represents the cost function defined on the set of edges E and p is the probability function defined on the set of edges E_N which satisfy condition (5). Note that $Z = Z^C \cup Z^N$, where Z^C is a subset of vertices of G which correspond to the set of controllable positions of dynamical system and Z^N is a subset of vertices of G which correspond to the set of uncontrollable positions of system L. In addition we shall use the notation Z_t^C and Z_t^N, where $Z_t^C = \big\{(x,t) \in Z_t | (x,t) \in Z^C\big\}$ and $Z_t^N = \big\{(x,t) \in Z_t | (x,t) \in Z^C\big\}$.

It is easy to observe that after the construction described above the problem 1 in the case $T_1 = T_2 = T$ can be formulated and solved on stochastic network $(G, Z^C, Z^N, z_0, z_f, p, T)$. A control $u(t)$ of system L in this network means a fixed passage from each controllable position $z = (x, t)$ to the next position $z = (x, t)$ through a leaving edge $e = (z, w) = \big((x, t), (y, t + 1)\big)$ generated by $u(t)$; this is equivalent with the prescription to these leaving edges the probability $p(e) = 1$ of the system's passage from the state (x, t) to the state $(y, t+1)$ considering $p(e) = 0$ for the rest of leaving edges. In other words a control on stochastic network means an extension of the probability function p from E_N to E by adding to the edges $e \in E \setminus E_N$ the probabilities $p(e)$ according to the mentioned above rule. We denote this probability function on E by p_u and will keep in mind that $p_u(e) = p(e)$ for $e \in E \setminus E_N$ and on E_C this function satisfies the following property

$$p_u : E_C \to \{0,1\}\,, \qquad \sum_{e \in E_C^+(z)} p_u(e) = 1 \text{ for } z \in Z^C$$

induced by the control $u(t)$ in the problems 1–6. If for the problems from section 2 the control $u(t)$ is given then we denote the stochastic network $(G, Z^C, Z^N, z_0, z_f, c, p_u, T)$; If the control $u(t)$ is not fixed then for the stochastic network we shall use the notation $(G, Z^C, Z^N, z_o, z_f, c, p, T)$. For the state probabilities of the system L on this stochastic network we shall use similar notations $P_{z_0}(u(t), z, T), P_{z_0}\big(u(t), z, T_1 \le T(z) \le T_2\big)$ and each time we will specify on which network they are calculated, i.e. will take into account that these probabilities are calculated by using the probability function on edges p_u which already do not depend on time.

Algorithm 1: *Determining the state probabilities of the system in Problem 1*

Preliminary step (Step 0): Put $P_{z_0}(u(t), z_0, 0) = 1$ for the position $z_0 \in Z$ and $P_{z_0}(u_p, z, t) = 0$ for the positions $z \in Z \setminus \{z_0\}$.

General step (Step $\tau, \tau \ge 1$): For every $z \in Z_\tau$ calculate

$$P_{z_0}(u(t), z, \tau) = \sum_{(w,z) \in E^-(z)} P_{z_0}(u(t), w, \tau - 1) p_u(w, z)$$

where $E^-(z) = \{(w, z) \in E \,|\, w \in Z_{\tau-1}\}$. If $\tau = T$ then stop; otherwise go to the next step.

The correctness of the algorithm follows from definition and network interpretation of the dynamics of system L. In the following we will consider that for the control problems from Section 2 the condition $U_t(x(t)) \neq \emptyset$ for every $x(t) \in X, t = 0, 1, 2, \dots, T_2 - 1$ holds.

Algorithm 2: *Determining the state probability of the system based on backward dynamic programming procedure*

Preliminary step (Step 0): Put $P_{z_f}(u(t), z_f, T) = 1$ for the position $z_f = (x_f, T)$ and $P_z(u(t), z, T) = 0$ for the the positions $z \in Z_T \setminus \{(x_f, T)\}$.

General step (Step $\tau, \tau \geq 1$): For every $z \in Z_{T-\tau}$ calculate

$$P_z(u(t), z_f, T) = \sum_{(z,w)\in E^+(z)} P_w(u(t), z_f, T) p_u(z, w)$$

where $E^+(z) = \{(z, w) \in E \,|\, w \in Z_{\tau+1}\}$. If $\tau = T$ then stop; otherwise go to next step.

Theorem 1. *For given control $u(t)$ Algorithm 2 correctly finds the state probabilities $P_{(x,T-\tau)}(u(t), x_f, T)$ for every $x \in X$ and $\tau = 0, 1, 2, \dots, T$. The running time of the algorithm is $O(|X|^2 T)$.*

Proof. The preliminary step of the algorithm is evident. The correctness of the general step of the algorithm follow from recursive formula at this general step which reflects dynamic programming principle for the state probabilities in simple stochastic process. In order to estimate the running time of the algorithm it is sufficient to estimate the number of elementary operations of general step of the algorithm. It is easy to see that the number of elementary operations for tabulation of state probabilities at the general step is $O(|X|^2)$. Taking into account that the number of steps of the algorithms is T we obtain that the running time of the algorithm is $O(|X|^2 L)$. □

Algorithm 3: *Determining the optimal control for Problem 1 with $T_1 = T_2 = T$*

We describe the algorithm for finding the optimal control $u^*(t)$ and the probabilities $P_{x(T-\tau)}(u^*(t), x_f, T)$ of system's passage from the states $x \in X$ at the moment of time $T - \tau$ to the state x_f by using τ units of time for $\tau = 0, 1, 2, \dots, T$. The algorithm consists of the preliminary, general and final steps. The preliminary and general steps of the algorithm find the values $\pi_{(x_f,T-\tau)}(z_f, T)$ of positions $(x, T-\tau) \in Z$ which correspond to probabilities $P_{x(T-\tau)}(u^*(t), z_f, T)$ of system passages from the state $x(T - \tau) \in X$ at the moment of time $T - \tau$ to the state $x_f(T) \in X$ at the moment of time T when the optimal control $u^*(t)$ is applied. At the end of the last iteration of general step of the algorithm 2 gives the subset of edges $E_C(u^*)$ of E_C which determines the optimal controls. The final step of the algorithm constructs an optimal control $u^*(t)$ of the problem.

Preliminary step (Step 0): Put $\pi_{z_f}(x_f, T) = 1$ for the position $z_f = (x_f, T)$ and $\pi_z(z_f, T) = 0$ for the positions $z \in Z_T \setminus \{(x_f, T)\}$; in addition put $E_C(u^*) = \emptyset$.

General step (Step $\tau \geq 1, \tau \geq 1$): For given τ do items a) and b):

a) For each uncontrollable position $z \in Z^N_\tau$ calculate

$$\pi_z(z_f, T) = \sum_{(z,w)\in E^+(z)} \pi_w(x_f, T) p(z, w);$$

b) For each controllable position $z \in Z_\tau^C$ calculate

$$\pi_z(z_f, T) = \max_{(z,w)\in E^+(x,T-\tau)} \pi_w(z_f, T)$$

and include in E_C^* edges (z, w) which satisfy the condition $\pi_z(z_f, T) = \pi_w(z_f, T)$. If $\tau = T$ then go to Final step; otherwise go to step $\tau + 1$.

Final Step: Form the graph $G^* = \big(Z, E_C^* \cup (E \setminus E_C)\big)$ and fix in G^* a map

$$u^* : (x, t) \to (y, t+1) \in X_{G^*}(x, t) \text{ for } (x, t) \in Z^C$$

where $X_{G^*} = \big\{(y, t+1) \in Z | \big((x, t), (y, t+1)\big) \in E_C^*\big\}$.

Theorem 2. *Algorithm 3 correctly finds the optimal control $u^*(t)$ and the state probability $P_{x(0)}(u^*(t), x_f, T)$ for an arbitrary starting position $x(0) \in X$ in problem 1 with $T = T_1 = T_2$. The running time of the algorithm is $O(|X|^2 T)$.*

Proof. The general step of the algorithm reflects the principle of optimality of dynamic programming for the problem of determining the control with maximal probabilities $P_{x(T-\tau)}(u^*(t), x_f, T) = \pi_{(x,T-\tau)}(x_f, T)$ of system's passages from the states $x \in X$ at the moment of time $T - \tau$ to the final state at the moment of time T. For each controllable position $(x, T - \tau) \in Z$ the values $\pi_{(x,T-\tau)}(z_f, T)$ are calculated on stochastic network in consideration that for given moment of time $T - \tau$ and given state $x \in X$ the optimal control $u^*(T - \tau) \in U_t(x(T - \tau))$ is applied. The computational complexity of the algorithm can be estimated in the same way as in Algorithm 2. Algorithm makes T steps and at each step uses $O(|X|^2)$ elementary operations. Therefore the running time of the algorithm is $O(|X|^2 T)$ □

4.2 Algorithm for determining the state probabilities in the case $T_1 \neq T_2$

We construct our network using the network $(G, Z^C, Z^N, z_0, z_f, c, p, T)$ with $T = T_2$ obtained according to the construction from Subsection 3.1. In this network we delete all edges originating in vertices (x, t) for $t = T_1, T_1+1, \ldots, T_2-1$ preserving edges originating in vertices (x, t) for $t = 0, 1, 2, \ldots, T_1-1$. We denote the stochastic network in this case $(G^0, Z^C, Z^N, z_0, Y, c, p, T_1, T_2)$, where $Y = \big\{(x_f, T_1), (x_f, T_1 + 1), \ldots, (x_f, T_2)\big\}$ and $G^0 = (Z, E^0)$ is the graph obtained from G dy deleting all edges which originate in vertices from Y, i.e $E^0 = E \setminus \{(z, w) \in E | z \in Y\}$. Let $P_{z_0}\big(u(t), Y, T_1 \le T(Y) \le T_2\big)$ be the probability of dynamical system to reach at least one of the states $(x_f, T_1), (x_f, T_1 + 1), \ldots, (x_f, T_2)$ at the moment of time t such that $T_1 \le t \le T_2$ if the dynamical system at the moment of time $\tau = 0$ has the state x_0.

Theorem 3. *For an arbitrary feasible control $u(t)$ and given staring state x_0 of dynamical system L the following formula holds*

$$P_{x_0}(u(t), x_f, T_1 \le T(x_f) \le T_2) = P_{z_0}(u(t), Y, T_1 \le T(Y) \le T_2). \qquad (10)$$

Proof. We prove the theorem by using induction principle on the number $k = T_2 - T_1$. Let us prove formula (10) for $k = 1$. In this case our network $(G^0, Z^C, Z^N, z_0, Y, c, p_u, T_1, T_2)$ is obtained from $(G, Z^C, Z^N, z_0, z_f, c, p, T_2)$ by deleting the edges $((x_f, T_1), (x_f, T_1 + 1))$ originating in (x_f, T_1). For this network we have $T_2 = T_2 + 1$ and $Y = (x_f, T_1), (x_f, T_1 + 1)$. Basing on formula (6) we can write the following equality

$$P_{x_0}(u(t), x_f, T_1 \leq T(x_f) \leq T_2) = P_{x_0}u(t), x_f, T_1) + P_{x_0}^{T_1, T_1+1}(u(t), x_f, T_1 + 1),$$

where $P_{x_0}^{T_1,T_1+1}(u(t), x_f, T_1 + 1)$ for given control $u(t)$ represents the probability of the system L to reach the state x_f from x_0 such that it does not pass at the moment of time T_1 through x_f. Taking into account that in our network all edges originating in (x_f, T_1) are deleted we obtain

$$P_{x_0}^{T_1,T_1+1}(u(t), x_f, T_1 + 1 = P_{z_0}(u(t), (x_f, T_1 + 1), T_1 + 1).$$

This means that

$$P_{x_0}(u(t), x_f, T_1 \leq T(x_f) \leq T_2) =$$
$$= P_{z_0}(u(t), (x_f, T_1), T_1) + P_{z_0}(u(t), (x_f, T_1 + 1), T_1 + 1).$$

If we use the property from Corollary 2 then we obtain formula (10) for $k = 1$.

Now assume that formula (10) holds for an arbitrary $k \geq 1$ an let us prove that it is true for $k + 1$.

We apply formula (7) for $P_{x_0}(u(t), x_f, T_1 \leq T(x) \leq T_2)$. Then we obtain

$$P_{x_o}(u(t), x_f, T_1 \leq T(x_f) \leq T_1 + k + 1) =$$
$$= P_{x_0}(u(t), x_f, T_1 \leq T(x_f) \leq T_1 + k) + P_{x_o}^{T_1,T_1+1,\dots,T_1+k}(u(t), x_f, T_1 + k + 1)$$

where$P_{x_o}^{T_1,T_1+1,\dots,T_1+k}(u(t), x_f, T_1 + k + 1)$ expresses the probability for the system L to reach the state x_f and it does not passes at the moment of time $T_1, T_1{+}1, \dots, T_1{+}k$ through the state x_f. According to the assumption of induction principle we can write

$$P_{x_o}(u(t), x_f, T_1 \leq T(x_f) \leq T_1 + k + 1) =$$
$$= P_{z_0}(u(t), Y \setminus (x_f, T_1 + k + 1), T_1 \leq T(Y \setminus (x, T_1 + k + 1) \leq T_1 + k) +$$
$$+ P_{x_o}^{T_1,T_1+1,\dots,T_1+k}(u(t), x_f, T_1 + k + 1).$$

Here in a similar way as in the case $k = 1$ holds

$$P_{x_o}^{T_1,T_1+1,\dots,T_1+k}(u(t), x_f, T_1 + k + 1) = P_{z_o}(u(t), (x_f, T_1 + k + 1), T_1 + k + 1)$$

because the stochastic network $(G^0, Z^C, Z^N, z_0, Y, c, p_u, T_1, T_2)$ is obtained from $(G, Z^C, Z^N, z_0, z_f, c, p, T_2)$ by deleting all edges originating in the vertices (x, T_1), $(x_f, T_1 + 1), \dots, (x_f, T_1 + k)$. So, the following formula holds

$$P_{x_o}(u(t), x_f, T_1 \leq T(x_f) \leq T_1 + k + 1) =$$

$$= P_{z_0}(u(t), Y \setminus (x_f, T_1+k+1), T_1 \leq T(Y \setminus (x_f, T_1+k+1) \leq T_1+k)+$$
$$+P_{z_o}(u(t), (x_f, T_1+k+1), T_1+k+1).$$

Now if we use the property from Corollary 2 of Lemma 1 then we obtain formula (10). □

Corollary 3. *For an arbitrary feasible control $u(t)$ and given staring state x_0 of dynamical system L the following formula holds*

$$P_{x_0}\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big) = \sum_{k=0}^{T_2-T_1} P_{z_0}\big(u(t), (x_f, T_1+k), T_1+k\big). \tag{11}$$

Basing on this result we can calculate $P_{x_0}\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$ in the following way. We apply Algorithm 1 on network $\big(G_f, Z^C, Z^N, z_0, Y, c, p_u, T_1, T_2\big)$ and determine the state probabilities $P_{z_0}(u(t), (x,\tau), \tau)$ for every $(x,\tau) \in Z$ and $\tau = 0, 1, 2, \ldots, T_2$. Then on the basis of formula (11) we find the probability $P_{x_o}\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$. We can use this fact for an another algorithm for finding the probability $P_x\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$. The algorithm finds the probabilities $P_z\big(u(t), Z_f, T_1 \leq T(Y) \leq T_2\big)$ on stochastic network $\big(G_f, Z^C, Z^N, z_0, Y, c, p_u, T_1, T_2\big)$ for every $z = (x, T-\tau) \in Z$. Then for $\tau = T$ we obtain $P_{x(T-\tau)}\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big) = P_{(x,T_2-\tau)}\big(u(t), Y, T_1 \leq T(Y) \leq T_2\big)$ for every $\tau = 0, 1, 2, \ldots, T_2$; if we fix $\tau = T_2$ then we find the probabilities $P_x\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$.

Algorithm 4: *Determining the solution of Problem 1 in the case $T_1 \neq T_2$*

Preliminary step (Step 0): Put $P_z\big(u(t), Y, T_1 \leq T(Y) \leq T_2\big) = 1$ for every position $z \in Y$ and $P_z\big(u(t), Y, T_1 \leq T(y) \leq T_2\big) = 0$ for the positions $z \in Z_{T_2} \setminus \{(x_f, T_2)\}$.

General step (Step $\tau, \tau \geq 1$): Calculate

$$P_z\big(u(t), Y, T_1 \leq T(Y) \leq T_2\big) = \sum_{(z,w)\in E^0(z)} P_w\big(u(t), Y, T_1 \leq T(Y) \leq T_2\big) p_u(z,w)$$

for every $z \in Z_{T_2-\tau} \setminus Y$ where $E^0(z) = \big\{(z,w) \in E^0 \mid w \in Z_{\tau+1}\big\}$. If $\tau = T$ then go to final step; otherwise go to step $\tau+1$.

Theorem 4. *Algorithm 4 correctly finds the state probability $P_{x(0)}(u^*(t), x_f, T)$ for an arbitrary starting position $x(0) \in X$ in problem 1 with $T_1 \leq T_2$. The running time of the algorithm is $O(|X|^2 T_2)$.*

Proof. In algorithm 4 the value $P_{x_0}\big(u(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$ is calculated on the basis of formula (11) applying Algorithm 2 for finding $P_{(x_0,0)}\big(u(t), (x_f, T_1+k), k\big)$ for $k = 0, 1, 2, \ldots, T_2 - T_1$. The application of Algorithm 2 on network with respect to each final position is equivalent with the specification of the preliminary step as it is described in Algorithm 4. So, the algorithm correctly finds the probability for the problem 1 with $T_1 \neq T_2$. The general step of the algorithm is made T_2 times. Therefore the running time of the algorithm is $O(|X|^2 T_2)$. □

Now let us show that the network $(G^0, Z^C, Z^N, z_0, Y, c, p_u, T_1, T_2)$ can be modified such that Algorithm 4 becomes Algorithm 2 on an auxiliary stochastic network. We make the following non-essential transformations of the structure of the network. In $G^0 = (Z, E^0)$ we add directed edges

$$\big((x_f, T_1), (x_f, T_1+1)\big), \big((x_f, T_1+1), (x_f, T_1+2)\big), \dots, \big((x_f, T_2-1), (x_f, T_2)\big).$$

To each directed edge $e_i = \big((x_f, T_1+i), (x_f, T_1+i+1)\big), i = 0, 1, 2, \dots, T_2 - T_1 - 1$ we define the values $p(e_i) = 1$ and $c(e_i) = 0$ which express respectively the probabilities and the costs of system's passage from the positions $(x_f, T_1 + i)$ to the position (x_f, T_1+i+1). We denote the network obtained after this construction by $(G^*, Z^C, Z^N, z_0, z_f, c^*, p_u^*, T_1, T_2)$, where $G^* = (Z, E^*)$ is the graph obtained from G^0 by using the construction described above, i.e. $E^* = E \cup \big\{\big((x_f, T_1+i), (x_f, T_1+i+1)\big),\quad i = 0, 1, 2, \dots, T_2 - T_1 - 1\big\}$; the probability and the cost functions $p_u^*, c^*$are obtained from p_u and c, respectively, according to given above additional construction. It is easy to see that if on this network we apply Algorithm 2 considering $T = T_2$ and $(x_f, T) = (x_f, T_2)$ then we find the state probabilities $P_{x_f, T_2-\tau}\big(u(t), (x_f, T_2), T_2\big)$ which coincide with the state probabilities $P_{x_f, T_2-\tau}\big(u(t), Y, T_1 \le T(Y) \le T_2\big)$.

Algorithm 5: *Determining the optimal control for Problem 2 with* $T_1 \neq T_1$

The algorithm consists of the preliminary, general and final steps. The preliminary and general steps find the values $\pi_{(x, T_2-\tau)}\big(Y, T_1 \le T(Y) \le T_2\big)$ which correspond to probabilities $P_{(x, T-\tau)}\big(u^*(t), Y, T_1 \le T(Y)\big) \le T_2$ when the optimal control is taken into account. So, these values represent the probabilities $P_{x(T-\tau)}\big(u^*(t), x_f, T_1 \le T(x_f) \le T_2\big)$ of system transactions from the states $x(T-\tau) \in X$ to the state x_f when the optimal control $u^*(t)$ is applied. At the end of the last iteration of general step the subset $E_C(u^*)$ from E_C is constructed. This subset determines the set of optimal controls for Problem 2. The final step of the algorithm fixes an optimal control $u^*(t)$.

Preliminary step (Step 0): Put $\pi_{(z,T)}(Y, T_1 \le T(Y) \le T_2) = 1$ for every position $z \in Y$ and $\pi_z(Y, T_1 \le T(Y) \le T_2) = 0$ for every positions $z \in Z_{T_2} \setminus \{(x_f, T_2)\}$; in addition put $E_C(u^*) = \emptyset$.

General step (Step $\tau, \tau \ge 1$): For given τ do the following items $a)$ and $b)$:

$a)$ For each position $z \in Z_\tau^N$ calculate

$$\pi_z\big(Y, T_1 \le T(Y) \le T_2\big) = \sum_{(z,w) \in E^+(z)} \pi_w\big(Y, T_1 \le T(Y) \le T_2\big) p(z, w);$$

$b)$ For each position $z \in Z_\tau^C$ calculate

$$\pi_z\big(Y, T_1 \le T(y) \le T_2\big) = \max_{(z,w) \in E^+(z)} \pi_w\big(Y, T_1 \le T(y) \le T_2\big)$$

and include in the set E_C^* each edge $e^* = (z, w)^*$ which satisfy the condition

$$\pi_z\big(Y, T_1 \le T(Y) \le T_2\big) = \pi_w\big(Y, T_1 \le T(Y) \le T_2\big).$$

If $\tau = T$ then go to Final step; otherwise go to step $\tau + 1$.

Final Step: Form the graph $G^* = \big(Z, E_C^* \cup (E \setminus E_C)\big)$ and fix in G^* a map

$$u^* : (x,t) \to (y, t+1) \in X_{G^*}(x,t) \quad \text{for} \quad (x,t) \in Z^C$$

where $X_{G^*} = \{(y, t+1) \in Z | ((x,t),(y,t+1)) \in E_C^*\}$.

Theorem 5. *Algorithm 5 correctly finds the optimal control $u^*(t)$ and the state probability $P_{x(0)}(u^*(t), x_f, T)$ for an arbitrary starting position $x(0) \in X$ in problem 1 with fixed final state $x_f \in X$ and given $T = T_1 = T_2$. The running time of the algorithm is $O(|X|^2T)$.*

Proof. The proof of this theorem is similar to the prove of Theorem 2. The general step of the algorithm reflects the principle of optimality of dynamic programming for the problem of finding the probabilities $P_{x(T-\tau)}\big(u^*(t), x_f, T_1 \leq T(x_f) \leq T_2\big)$. These probabilities in stochastic networks correspond to the probabilities $P_{(x,T_2-\tau)}\big(u^*(t), Y, T_1 \leq T(Y) \leq T_2\big) = \pi_{(x,T_2-\tau)}(Y, T_1 \leq T(Y) \leq T_2)$. For each controllable position $(x, T-\tau)$ the values $\pi_{(x,T_2-\tau)}\big(Y, T_1 \leq T(Y) \leq T_2\big)$ are calculated in consideration that for given moment of time $T - \tau$ and given state x the optimal control $u^*(T_2 - \tau) \in U_t(x(T_2 - \tau))$ is applied. Therefore $\pi_{(x,T_2-\tau)}(Y, T_1 \leq T(Y) \leq T_2) = P_{x(T_2-\tau)}\big(u^*(t), x_f, T_1 \leq T(Y) \leq T_2\big)$ for every $x \in X$ and $\tau = 0, 1, 2, \dots, T_2$. Taking into account that at each step the directed edges e^* correspond to the optimal control for the corresponding positions on stochastic network, we obtain at the final step the set of edges E^* which give the optimal control for arbitrary state x and arbitrary moment of time t. In the same way as in previous algorithms we can show that the running time of the algorithm is $O(|X|^2T_2)$. □

5 Algorithms for Determining the Expectation of Integral-Time Cost in Problems 3–6

In this section we describe algorithms for calculation of the expected integral-time costs of state transactions of dynamical system in problems 3-6.

5.1 Calculation of the Expectation of Integral-Time cost in Problem 3

The expectation of integral-time cost for dynamical system L on stochastic network $(G, Z^C, Z^N, z_o, c, p_u, T)$ in problem 3 is defined in analogues way as in Subsection 3.2 using the following recursive formula:

$$C_z(u(t), T) = \sum_{(z,w) \in E^+(z)} p_u(z,w)\big(c(z,w) + C_w(u(t), T)\big),$$

$$z \in Z_{T-\tau}, \tau = 1, 2, \dots, T,$$

where $E^+(z) = \{(z,w) \in E | w \in Z_{T-\tau+1}\}$. This formula can be treated in the following way. Assume that we should estimate the expected integral-time cost of system's transactions during τ units of time when the system starts transactions in position $z = (x, T-\tau)$ at the moment of time $T-\tau$. If the system makes a transition from the position $z = (x, T-\tau)$ to the position $w = (y, T-\tau+1)$ it will spend the amount $c(z, w)$ plus the amount it expects to spend if the system starts the remained $\tau - 1$ transactions in the position $w = (y, T-\tau+1)$. Therefore if the system L at the moment of time $T-\tau$ is in position $z = (x, T-\tau)$ then the expected integral-cost of system's transitions from z must be weighted by he probabilities of such transactions $p_u(z, w)$ to obtain the total expected integral-time costs.

Algorithm 6: *Determining the expectation of integral-time cost in Problem 3*
Preliminary Step (Step 0): Put $C_z(u(t), T) = 0$ for every $z \in Z_T$.
General Step (Step τ,$\tau \geq 1$): For each $z \in Z_{T-\tau}$ calculate

$$C_z(u(t), T) = \sum_{(z,w)\in E^+(z)} p_u(z, w)\big(c(z, w) + C_w(u(t), T)\big).$$

If $\tau = T$ then stop; otherwise go to step $\tau + 1$.

Algorithm 6 uses the backward dynamic procedure and finds $C_z(u(t), T)$ for every position $z \in Z$. For a fixed position $z = (x, T-\tau) \in Z$ the value $C_z(u(t), T)$ corresponds to the expected integral-time cost $C_{x(T-\tau)}(u(t), T)$ of the system in the next τ transactions when it starts in the state $x = x(T-\tau)$ at the moment of time $T-\tau$, i.e. $C_{(x,T-\tau)}(u(t), T) = C_{x(o)}(u(t), T)$.

Algorithm 7: *Determining the optimal control for problem 4*

The algorithm consists of the preliminary, general and final steps. At the preliminary and general steps the algorithm finds the optimal values of the expectation of integral-time costs $C_z(u(t), T)$ which in algorithm are denoted by $Exp_z(T)$. For a position $z = (x, T-\tau)$ the value $Exp_z(T)$ expresses the expected integral-time cost during τ transactions of the system when it starts transactions in the state $x = x(T-\tau)$ at the moment of time $T-\tau$.This value is calculated in the consideration that the optimal control $u(t)$ is applied. In addition at the general step of the algorithm the possible directed edges $e^* = \big((x, T-\tau), (y, T-\tau+1)\big)^*$ which correspond to optimal control in the state $x = x(T-\tau)$ at the moment of time $T-\tau$ are cumulated in the set $E_C(u^*)$. The set of optimal controls is determined by $E_C(u^*)$; at the final step an optimal control is fixed.

Preliminary step (Step 0): Put $Exp_z(T) = 0$ for $z \in Z_T$ and $E_C(u) = \emptyset$.
General step (Step τ, $\tau \geq 1$): For given τ do the following items a) and b):
a) For each uncontrollable position $z \in Z^N_{T-\tau}$ calculate

$$Exp_z(T) = \sum_{(z,w)\in E^+(z)} p(z, w)\big(c(z, w) + Exp_w(T)\big);$$

b) For each controllable position $z \in Z^C_{T-\tau}$ calculate

$$Exp_z(T) = \max_{(z,w)\in E^+(z)} \big(c(z, w) + Exp_w(T)\big)$$

and include in the set E_C^* each edge $e^* = (z,w)^*$ which satisfies the condition

$$c((z,w)^*) + Exp_{w^*}(T) = \max_{(z,w)\in E^+(z)} \big(c(z,w) + Exp_w(T)\big).$$

If $\tau = T$ then go to Final step; otherwise go to step $\tau + 1$.

Final Step: Form the graph $G^* = (Z, E_C^* \cup (E \setminus E_C))$ and fix in G^* a map $u^* : (x,t) \to (y, t+1) \in X_{G^*}(x,t) \quad for \quad (x,t) \in Z^C$.

Theorem 6. *Algorithm 7 correctly finds the optimal control $u^*(t)$ and the expected integral-time costs $C_{x(0)}(T)$ of the system's transactions during T units of time from an arbitrary starting position $x = x(0) \in X$ in problem 4. The running time of the algorithm is $O(|X|^2 T)$.*

This theorem can be proved in analogues way as Theorem 2.

5.2 Determining the Expectations of Integral-Time cost in Problems 5 and 6

For problems 5 and 6 we need to precise what is meant by the expectation of integral-time cost for dynamical system when the state x_f is reached at the moment of time $T(x)$ such that $T_1 \le T(x) \le T_2$. At first let us analyze the case $T_1 = T_2 = T$. We consider this problem on stochastic network $(G, Z^C, Z^N, z_0, z_f, c, p_u, T)$. If we assume that the final position $z_f = (x_f, T)$ is reached at the moment of time T then we should consider that the probability of system transaction from an arbitrary starting position $z = (x, 0)$ to the position z_f is equal to 1. This means that the probabilities $p_u(e)$ on edges $e \in E$ should be redefined or transformed in such way that the mentioned above condition on stochastic network holds.We denote these redefined values by $p'_u(e)$ and call them conditional probabilities. It is evident that if the system never can meet a directed edge $e \in E$ during transition from a position $(x, 0)$ to the position z_f then the conditional probability $p'_u(e)$ of this edge is equal to zero. So, the first step we should do in the transformation is to delete all such edges from the graph G. After such transformation we obtain a new graph $G' = (Z, E')$ in which for some positions $z \in Z$ the condition $\sum_{(z,w)\in E'(z)} p(z,w) = 1$ is not satisfied (here $E'(z)$ represents the subset of edges from E' which in vertex z, i.e. $E'(z) = \{(z,w) | (z,w) \in E'\}$). Then we find for each position $z \in Z$ the value $\pi(z) = \sum_{(z,w)\in E'(z)} p_u(z,w)$ and after that for an arbitrary position $z \in Z$ with $\pi(z) \neq 0$ we make the transformation

$$p'_u(z,w) = \frac{1}{\pi(z)} p_u(z,w)$$

for every $(z,w) \in E'(z)$. After these transformations we can apply Algorithm 6 on stochastic network $(G, Z^C, Z^N, z_0, c, p'_u, T)$ with conditional probabilities $p'_u(e)$ of edges $e \in E$; here $p'_u(e) = 0$ for $e \in E \setminus E'$. If for this network we find $C_{z_0}(T)$

then fix $C_{z_0}(T) = C_{x_0}(u(t), x_f, T) = C_{z_0}(T)$, i. e. this value represents the expected integral-time cost of dynamical system L in problem 5. In the case $T_1 \neq T_2$ the expected integral-time cost $C_{x_o}(u(t), x_f, T_1 \leq T(x) \leq T_2)$ can be found in analogues way if we consider problem 5 on stochastic network $(G^*, Z^C, Z^N, z_0, z_f, c^*, p^*, T_1, T_2)$ and will make a similar transformation. It is evident that the control problem 6 can be reduced to control problem 4 using the approach described above. This allows us to find the optimal control $u^*(t)$ which provides a maximal expected integral-time cost $C_{x_0}(u^*(t), x_f, T_1 \leq T(x_f) \leq T_2)$ of system transactions from starting state x_0 to final state x_f such that $T_1 \leq T(x_f) \leq T_2$.

References

[1] Bellman R., Kalaba R. *Dynamic programming and modern control theory.* Academic Press, New York and London (1965).

[2] Boltjanski W.G. *Optimale Steuerung diskreter Systeme.* Leipzig Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1976.

[3] Feller W. *An Introduction to Probability Theory and its Applications*, vol. 1, New York, John Wiley, 2nd edition, 1957.

[4] Howard R.A. *Dynamic Programming and Markov Processes.* Wiley, 1960.

[5] Lozovanu D., Pickl S. *Optimization and Multiobjective Control of Time-Discrete Systems.* Springer, 2008.

[6] Lozovanu D., Pickl S. *Allgorithms and the calculation of Nash equilibria for multi-objective control of time-discrete systems and polynomial-time algorithms for dynamic c-games on networks.* European Journal of Operational Research, 2007, **181**, 1214–1232.

Dmitrii Lozovanu
Institute of Mathematics and Computer Science
Academy of Sciences of Moldova
5 Academiei str. Chişinău, MD−2028
Moldova
E-mail: *lozovanu@math.md*

Received January 25, 2009

Stefan Pickl
Institut fur Angewandte Systemwissenschaften und
Wirtschaftsinformatik
Fakultät fur Informatik, Universität der Bundeswehr,
München
E-mail: *stefan.pickl@unibw.de*