Chapter 2
Stochastic Optimal Control Problems and Markov Decision Processes with Infinite Time Horizon

The aim of this chapter is to develop methods and algorithms for determining the optimal solutions of stochastic discrete control problems and Markov decision problems with an infinite time horizon. We denote such methods and algorithms on the bases of the results from the previous chapter and classical optimization methods. The set of states of the system in the considered problems is finite and the starting state is fixed. We study the stochastic discrete processes that may be controlled in some dynamical states. The average and the expected total discounted costs optimization principles for such processes are applied and new classes of a stochastic control model are formulated. Based on such a concept we study a class of stochastic discrete control problems that emphasis Markov decision problems and deterministic optimal control problems with an infinite time horizon. We obtain the stochastic versions of classical discrete control problems assuming that the dynamical system in the control process may admit dynamical states in which the vector of control parameters is changing in a random way according to given distribution functions of the probabilities on given feasible sets. So, in the considered control problems we assume that the dynamics of the system may contain controllable states as well as uncontrollable states. These problems are formulated on networks and polynomial time algorithms for determining their optimal solutions are proposed. In the case that the dynamical system contains only controllable states the proposed algorithms become algorithms for determining the optimal stationary strategies of the classical deterministic control problems with an infinite time horizon. The proposed methods and algorithms are extended to Markov decision processes.

We develop a linear programming approach to Markov decision processes and show how to use the duality theory for determining solutions of the decision problems with average and expected total discounted optimization criteria. Based on such an approach we describe algorithms for solving new classes of stochastic discrete optimization problems. Polynomial time algorithms for Markov decision problems with average and expected total discounted costs optimization criteria are proposed and formulated.

Furthermore, some numerical examples are given and the computational complexity aspects of the described methods and algorithms are analyzed.

D. Lozovanu and S. Pickl, *Optimization of Stochastic Discrete Systems and Control on Complex Networks*, Advances in Computational Management Science 12,
DOI 10.1007/978-3-319-11833-8_2

2.1 Problem Formulation and the Main Concept of Optimal Control Models with Infinite Time Horizon

The infinite horizon decision problem can be regarded as approximation model for decision problems with an finite time horizon in the case of a large sequence of decisions. Often, it is easier to solve the infinite horizon problem and to use the solution of this to obtain a solution of the finite horizon problem with a large number of decisions. The Markov decision processes and the classical control problems with infinite time horizon are related to such kind of models that are widely used for studying and solving many practical finite horizon decision problems. Below we formulate a class of stochastic discrete optimal control problems with average and expected total discounted costs optimization criteria that combine the statements of deterministic optimal control problems with infinite time horizon and Markov decision processes [5, 114]. We start with a formulation of the stochastic optimal control problem that represents a generalization of the following deterministic control model.

Let a discrete dynamical system $\mathbb{L}$ with a finite set of states $X \subset \mathbb{R}^n$ be given where at every time-step $t = 0, 1, 2, \ldots$, the state of the system $\mathbb{L}$ is $x(t) \in X$. At the starting moment of time $t = 0$ the state of the dynamical system $\mathbb{L}$ is $x(0) = x_0$. Assume that the dynamics of the system $\mathbb{L}$ is described by the system of difference equations

$$x(t+1) = g_t(x(t), u(t)), \quad t = 0, 1, 2, \ldots \tag{2.1}$$

where

$$x(0) = x_0 \tag{2.2}$$

and

$$u(t) = (u_1(t), u_2(t), \ldots, u_m(t)) \in \mathbb{R}^m$$

represents the *vector of the control parameters* (see [6, 11, 132]). For any time step t and an arbitrary state $x(t) \in X$ the feasible set $U_t(x(t))$ of the vector $u(t)$ of control parameters is given, i.e.,

$$u(t) \in U_t(x(t)), \quad t = 0, 1, 2, \ldots. \tag{2.3}$$

We assume that in (2.1) the vector functions

$$g_t(x(t), u(t)) = (g_t^1(x(t), u(t)),\ g_t^2(x(t), u(t)), \ldots,\ g_t^n(x(t), u(t)))$$

are determined uniquely by $x(t)$ and $u(t)$ at every time step $t = 0, 1, 2, \ldots$. So, $x(t+1)$ is determined uniquely by $x(t)$ and $u(t)$.

Additionally, we assume that at each moment of time t the cost

$$c_t(x(t), x(t+1)) = c_t(x(t),\ g_t(x(t), u(t)))$$

of the system's transition from the state $x(t)$ to the state $x(t+1)$ is known.

Let

$$x_0 = x(0), x(1), x(2), \ldots, x(t), \ldots$$

be a trajectory generated by given vectors of the control parameters

$$u(0), u(1), \ldots, u(t-1), \ldots .$$

Then after a fixed number of transitions τ of the dynamical system we can calculate the *integral-time cost (total cost)* which we denote by $F^{\tau}_{x_0}(u(t))$, i.e.,

$$F^{\tau}_{x_0}(u(t)) = \sum_{t=0}^{\tau-1} c_t(x(t),\ g_t(x(t), u(t))). \tag{2.4}$$

In [6, 11] the following discrete optimal control problem with finite time horizon has been considered: Find for given τ the vectors of control parameters

$$u(0), u(1), u(2), \ldots, u(\tau-1)$$

which satisfy the conditions (2.1)–(2.3) and minimize the functional (2.4). The solution of this optimal control problem can be found by using *dynamic programming techniques* [6, 79].

Here we consider the *infinite horizon control model*. We assume that τ is not bounded, i.e., $\tau \to \infty$. It is evident that if $\tau \to \infty$ then the integral-time cost

$$\lim_{\tau\to\infty} \sum_{t=0}^{\tau-1} c_t(x(t),\ g_t(x(t), u(t)))$$

for a given control may not exist. Therefore, we study in this case the asymptotic behavior of the integral-time cost $F^{\tau}_{x_0}(u(t))$ by a trajectory determined by a feasible or an optimal control. To estimate this value we apply the concept from [5, 6], i.e., for a fixed control u if τ is too large we estimate $F^{\tau}_{x_0}(u(t))$ asymptotically using the function $\phi_u(\tau) = K\varphi(\tau)$ such that

$$\lim_{\tau\to\infty} \frac{1}{\varphi(\tau)} \sum_{t=0}^{\tau-1} c_t(x(t),\ g_t(x(t), u(t))) = K, \tag{2.5}$$

where K is a constant.

So, in control problems with an infinite time horizon we are seeking for a control u^* with a suitable limiting function $\phi_{u^*}(\tau)$.

Based on the asymptotic approach mentioned above we may conclude that for a given control, if τ is too large, the value $F^{\tau}_{x_0}(u(t))$ can be approximated by $K\varphi(\tau)$.

Moreover, we can see that for the stationary case of the control model with the costs that do not depend on time the function $\phi_u(\tau)$ is linear. This means that $\varphi(\tau) = \tau$ and $F^{\tau}_{x_0}(u(t))$ for a large τ can be approximated by $\phi_u(\tau) = K\tau$.

In the following we study only stationary control problems. For such problems the vector functions g_t and the feasible sets $U_t(x(t))$ do not depend on time, i.e., $g_t(x,u) = g(x,u)$ and $U_t(x) = U(x)$, $\forall x \in X$, $t = 0, 1, 2, \ldots$. Moreover, the control at every discrete moment of time depends only on the state $x \in X$ and the cost of the system's transition from the state $x \in X$ to the state $y \in Y$ does not depend on time, i.e., $c_t(x(t), x(t+1)) = c(x,y)$, $\forall x, y \in X$ and every $t = 0, 1, 2, \ldots$ if $x = x(t)$, $y = x(t+1)$.

Thus, for the considered stationary control problems the integral-time cost by a trajectory during τ transitions can be asymptotically expressed as $F^{\tau}_{x_0}(u(t)) = K\varphi(\tau)$, where $\varphi(\tau) = \tau$. In this case for the dynamical system $\mathbb{L}$ the constant K in (2.5) expresses the *average cost per transition* along a trajectory determined by the control $u(t)$. Therefore, for the infinite horizon optimal control problem the objective function which has to be minimized is defined as follows:

$$F_{x_0}(u(t)) = \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=0}^{\tau-1} c(x(t),\ g(x(t), u(t))). \tag{2.6}$$

In [5] it is shown that for the stationary case of the problem the optimal control u^* does not depend on time or on the starting state and it can be found in the set of stationary controls.

Another class of control problems with an infinite time horizon which is widely used for practical problems is characterized by a discounting objective cost function [8]

$$\widehat{F}_{x_0}(u(t)) = \sum_{t=0}^{\infty} \gamma^t c_t(x(t),\ g_t(x(t), u(t))). \tag{2.7}$$

Here γ is a *discount factor* that satisfies the condition $0 < \gamma < 1$ and $\widehat{F}_{x_0}(u(t))$ is called the *total discounted cost*. In a control problem with such an optimization criterion we are seeking for the control which minimizes the functional (2.7).

In [28, 114, 129, 140] it is shown that if $0 < \gamma < 1$ and the costs $c_t(x(t), g_t(x(t), u(t)))$ are bounded then for the stationary case of the control problem with a discounted objective optimization criterion the optimal stationary control exists.

The problems formulated above correspond to deterministic models in which the decision maker is able to fix the vector of control parameters $u(t)$ from a given feasible set $U_t(x(t))$ in each dynamical state $x(t)$; the states $x(t) \in X$ in these models are called *controllable states*.

The main results we describe in the following are related to stochastic versions of the control problems formulated above. We consider the control models in which the dynamical system in the control process may admit dynamical states $x(t)$ where the corresponding vector of control parameters $u(t)$ is changed in a random way according to given distribution functions

$$p : U_t(x(t)) \to [0,1], \quad \sum_{i=1}^{k(x(t))} p\left(u^i_{x(t)}\right) = 1 \tag{2.8}$$

on the corresponding dynamical feasible set $U_t(x(t))$. Here $k(x(t)) = |U_t(x(t))|$, i.e., we consider the control models with finite feasible sets.

We regard each dynamical state $x(t)$ of the system in the considered control problem as a position (x, t) and we assume that the set of positions

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

is divided into two subsets

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

such that Z^C corresponds to the set of *controllable states* and Z^N corresponds to the set of *uncontrollable states*. This means that for the stochastic control problems we have the following behavior of the dynamics in the control process: If the starting state $x(0)$ belongs to the set of controllable states Z^C then the decision maker fixes the vector of control parameters $u(0)$ from the feasible set $U_0(x(0))$ and we obtain the next state $x(1)$; if the state $x(0)$ belongs to the set Z^N then the system passes to the next state $x(1)$ in a random way. If at the moment of time $t = 1$ the state $x(1)$ belongs to the set of controllable states Z^C then the decision maker fixes the vector of control parameters $u(1)$ from $U_1(x(1))$ and we obtain the next state $x(2)$; if $x(1)$ belongs to the set of uncontrollable states Z^N then the system passes to the next state $x(2)$ in a random way and so on indefinitely.

It is evident that for a fixed control the average cost per transition and the discounted total cost in this process represent the random variables induced by the distribution functions on feasible sets in the uncontrollable states and the control in the controllable states.

To define the *expected average cost per transition* and *expected discounted total cost* in the considered stochastic control problems for a fixed control we will apply the concept of Markov decision processes in the following way:

Let $u'(t) \in U_t(x(t))$ be the given feasible vectors in the controllable states $x(t) \in Z^C$. Then we may assume that we have the following distribution functions

$$p : U_t(x(t)) \to \{0, 1\} \quad \text{for } x(t) \in Z^C$$

where $p(u'(t)) = 1$ and $p(u(t)) = 0$, $\forall u(t) \in U_t(x(t))\backslash\{u'(t)\}$.

These distribution functions in the controllable states together with the distribution functions (2.8) in the uncontrollable states determine a Markov process. For this Markov process with transition probabilities $p_{z,v}$ and transition costs $c_{z,v}$ for $(z, v) \in Z \times Z$ we can determine the expected average and the expected discounted total costs which we denote, respectively, by $F_{x_0}(u(t))$ and $\widehat{F}_{x_0}(u(t))$. In such a way we obtain the corresponding optimization problems in which we are seeking for the controls that minimize the expected average and discounted total costs, respectively.

Thus, we shall use the combined concept of deterministic and stochastic control models from [36, 81–94, 96, 108, 109], and will develop algorithms for determining optimal strategies of the considered problems. Mainly, we will study the stationary

versions of the control problems with a finite set of states for the dynamic system and will describe algorithms based on linear programming. In the general case, for non-stationary control problems, the optimal control may not exist. Some special classes of non-stationary problems may admit the solution and the optimal control can be found by using a special calculation procedure.

2.2 An Optimal Stationary Control with an Average Cost Criterion and Algorithms for Solving Stochastic Control Problems on Networks

In this section we consider the stationary stochastic discrete optimal control problem with average cost criterion. We formulate this problem on networks and describe polynomial time algorithms for determining the optimal control by using a linear programming approach.

2.2.1 Problem Formulation

Let a discrete dynamical system $\mathbb{L}$ with a finite set of states X be given, where $|X| = n$. At every discrete moment of time $t = 0, 1, 2, \ldots$ the state of $\mathbb{L}$ is $x(t) \in X$. 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 and an arbitrary directed edge $e = (x, y) \in E$ expresses the possibility of the system $\mathbb{L}$ to pass from the state $x = x(t)$ to the state $y = x(t + 1)$ at every discrete moment of time t. So, a directed edge $e = (x, y)$ in G corresponds to a stationary control of the system in the state $x \in X$ which provides a transition from $x = x(t)$ to $y = x(t + 1)$ for every discrete moment of time t. We assume that graph G does not contain deadlock vertices, i.e., for each x there exists at least one leaving directed edge $e = (x, y) \in E$. In addition, we assume that to each edge $e = (x, y) \in E$ a quantity c_e is associated which expresses the cost (or the reward [47]) of the system $\mathbb{L}$ to pass from the state $x = x(t)$ to the state $y = x(t)$ for every $t = 0, 1, 2, \ldots$.

The cost c_e for an arbitrary edge $e = (x, y)$ is denoted by $c_{x,y}$. A sequence of directed edges $E' = \{e_0, e_1, e_2, \ldots, e_t, \ldots\}$ where $e_t = (x(t), x(t + 1))$, $t = 0, 1, 2, \ldots$ determines in G a control of the dynamical system with a fixed starting state $x_0 = x(0)$. An arbitrary control in G generates a trajectory $x_0 = x(0), x(1), x(2), \ldots$ for which the average cost per transition can be defined in the following way

$$f(E') = \lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} c_{e_\tau}.$$

In [5] it is shown that this value exists and $|f_{x_0}(E')| \leq \max_{e \in E'} |c_e|$. Moreover, in [5] it is shown that if G is strongly connected then for an arbitrary fixed starting state $x_0 = x(0)$ there exists the optimal control $E^* = \{e_0^*, e_1^*, e_2^* \ldots\}$ for which

$$f(E^*) = \min_{E'} \lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} c_{e_\tau}$$

and this optimal control does not depend either on the starting state or on time. Therefore, the optimal control for this problem can be found in the set of stationary strategies $\mathbb{S}$. A *stationary strategy* in G is defined as a map:

$$s : x \to y \in X(x) \quad \text{for } x \in X,$$

where $X(x) = \{y \in X \,\big|\, e = (x, y) \in E\}$.

Let s be a stationary strategy. Denote by $G_s = (X, E_s)$ the subgraph of G generated by edges of the form $e = (x, s(x))$ for $x \in X$. Then it is easy to observe that in G_s there exists a unique directed cycle C_s which can be reached from x_0 through the directed edges from E_s. Moreover, we can see that the mean cost of this cycle is equal to the average cost per transition of the dynamical system by the trajectory generated by the stationary strategy s. Thus, if G is a strongly connected directed graph then the problem of determining the optimal control on G is equivalent to the problem of finding in G the cycle C_G^* for which

$$\frac{\sum_{e \in E(C_G^*)} c_e}{n(C_G^*)} = \min_{C_G} \frac{\sum_{e \in E(C_G)} c_e}{n(C_G)},$$

where $E(C_G)$ is the set of directed edges of the directed cycle C_G in G that can be reached from a starting vertex and $n(C_G)$ is the number of its edges. If the cycle C_G^* is known then the optimal control for an given arbitrary starting state $x_0 = x(0)$ in G can be found in the following way: We fix the transitions through the directed edges of the graph in order to reach a vertex of the directed cycle C_G^* and then we preserve transitions through the directed edges of this cycle.

Polynomial and strongly polynomial time algorithms for determining the optimal average cost cycles in a weighted directed graph and the optimal stationary strategies for control problems on networks have already been proposed in [53, 65, 79, 117].

In the following we will consider the stochastic version of the problem formulated above. We assume that the set of states X of the dynamical system may admit states in which the system $\mathbb{L}$ makes transitions to the next state in a random way according to a given distribution function of probabilities on the set of possible transitions from these states. So, the set of states X is divided into two subsets X_C and X_N ($X = X_C \cup X_N$, $X_C \cap X_N = \emptyset$), where X_C represents the set of states $x \in X$ in which the transitions of the system to the next state y can be controlled by the decision maker at every discrete moment of time t and X_N represents the set of states $x \in X$ in which the decision maker is not able to control the transition because the

system passes to the next state y randomly. Thus, for each $x \in X_N$ a probability distribution function $p_{x,y}$ on the set of possible transitions (x, y) from x to $y \in X(x)$ is given, i.e.,

$$\sum_{y \in X(x)} p_{x,y} = 1, \quad \forall x \in X_N; \quad p_{x,y} \geq 0, \quad \forall y \in X(x). \tag{2.9}$$

Here $p_{x,y}$ expresses the probability of the system's transition from the state x to the state y for every discrete moment of time t. Note, that the condition $p_{x,y} = 0$ for a directed edge $e = (x, y) \in E$ is equivalent with the condition that G does not contain this edge.

In the same way as for the deterministic problem here we assume that to each directed edge $e = (x, y) \in E$ a cost c_e is associated.

We call the graph G with the properties mentioned above *decision network* and denote it by (G, X_C, X_N, c, p, x_0). So, this network is determined by the directed graph G with a fixed starting state x_0, the subsets X_C, X_N, the cost function $c : E \to \mathbb{R}$ and the probability function $p : E_N \to [0, 1]$ on the subset of the edges $E_N = \{e = (x, y) \in E \,\big|\, x \in X_N, \ y \in X\}$ where p satisfies the condition (2.9). If the control problem is considered for an arbitrary starting state then we denote the network by (G, X_C, X_N, c, p).

We define a stationary strategy for the control problem on networks as a map:

$$s : x \to y \in X(x) \quad \text{for} \quad x \in X_C.$$

Let s be an arbitrary stationary strategy. Then we can determine the graph $G_s = (X, E_s \cup E_N)$, where $E_s = \{e = (x, y) \in E \,|\, x \in X_C, \ y = s(x)\}$, $E_N = \{e = (x, y) \,|\, x \in X_N, \ y \in X\}$. This graph corresponds to a Markov process with the probability matrix $P^s = (p^s_{x,y})$, where

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

In the considered Markov process for an arbitrary state $x \in X_C$ the transition $(x, s(x))$ from the states $x \in X_C$ to the states $y = s(x) \in X$ is made with the probability $p_{x,s(x)} = 1$ if the strategy s is applied. For this Markov process we can determine the average cost per transition for an arbitrary fixed starting state $x_i \in X$ in such a way as we have defined it in Sect. 1.7.2. Thus, we can determine the vector of average costs ω^s which corresponds to the strategy s. As we have shown in Sect. 1.7.2 the vector ω^s can be calculated according to the formula $\omega^s = Q^s \mu^s$, where Q^s is the limit matrix of the Markov process generated by the stationary strategy s and μ^s is the corresponding vector of the immediate costs, i.e., $\mu^s_x = \sum_{y \in X(x)} p^s_{x,y} c^s_{x,y}$. A component ω^s_x of the vector ω^s represents the average cost per transition in our problem with a given starting state x and a fixed strategy s, i.e.,

$$f_x(s) = \omega^s_x.$$

In such a way we can define the value of the objective function $f_{x_0}(s)$ for the control problem on a network with a given starting state x_0 when the stationary strategy s is applied.

The control problem on the network (G, X_C, X_N, c, p, x_0) consists of finding a stationary strategy s^* for which

$$f_{x_0}(s^*) = \min_s f_{x_0}(s).$$

In the next section we can see that the optimal stationary strategy in the considered problem does not depend on the starting state. We show that a polynomial time algorithm for determining the optimal solution of this problem can be elaborated. Moreover, we show that the proposed algorithm can be extended to Markov decision processes.

2.2.2 A Linear Programming Approach for Determining Optimal Stationary Strategies on Perfect Networks

We consider the stochastic control problem on the network (G, X_C, X_N, c, p, x_0) with $X_C \neq \emptyset$, $X_N \neq \emptyset$ and assume that G is a strongly connected directed graph. Additionally, we assume that in G for an arbitrary stationary strategy $s \in \mathbb{S}$ the subgraph $G_s = (X, E_s \cup E_N)$ is strongly connected. This means that the Markov chain induced by the probability transition matrix P^s is irreducible for an arbitrary strategy s. We call the decision network with such a condition a *perfect network*. At first we describe an algorithm for determining the optimal stationary strategies for the control problem on perfect networks. Then we show that the proposed algorithm can be extended for the problem if an arbitrary strategy s generates a Markov unichain. For a unichain control problem the graph G^s induced by a stationary strategy may not be strongly connected but it contains a unique strongly connected component that is reachable from every $x \in X$.

So, in this section we consider the control problem that the average cost per transition is the same for an arbitrary starting state, i.e.,

$$f_x(s) = \omega^s, \quad \forall x \in X.$$

We will consider in the next section the case of a multichain control problem, i.e., the case that for different starting states the average cost per transition may be different.

Let $s \in \mathbb{S}$ be an arbitrary strategy. Taking into account that for every fixed $x \in X_C$ we have a unique $y = s(x) \in X(x)$ then we can identify the map s with the set of boolean values $s_{x,y}$ for $x \in X_C$ and $y \in X(x)$, where

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

For the optimal stationary strategy s^* we denote the corresponding boolean values by $s^*_{x,y}$.

Assume that the network (G, X_C, X_N, c, p, x_0) is perfect. Then the following lemma holds.

Lemma 2.1 *A stationary strategy s^* is optimal if and only if it corresponds to an optimal solution q^*, s^* of the following mixed integer bilinear programming problem: Minimize*

$$\psi(s, q) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y} s_{x,y} q_x + \sum_{z \in X_N} \mu_z q_z \tag{2.10}$$

subject to

$$\begin{cases} \sum\limits_{x \in X_C} s_{x,y} q_x + \sum\limits_{z \in X_N} p_{z,y} q_z = q_y, \quad \forall y \in X; \\ \sum\limits_{x \in X_C} q_x + \sum\limits_{z \in X_N} q_z = 1; \\ \sum\limits_{y \in X(x)} s_{x,y} = 1, \quad \forall x \in X_C; \\ s_{x,y} \in \{0, 1\}, \ \forall x \in X_C, \ y \in X; \quad q_x \geq 0, \ \forall x \in X, \end{cases} \tag{2.11}$$

where

$$\mu_z = \sum_{y \in X(z)} p_{z,y} c_{z,y}, \quad \forall z \in X_N.$$

Proof Denote $\mu_x = \sum_{y \in X(x)} c_{x,y} s_{x,y}$ for $x \in X_C$. Then μ_x for $x \in X_C$ and μ_z for $z \in X_N$ represent, respectively, the immediate cost of the system in the states $x \in X_C$ and $z \in X_N$ if the strategy $s \in S$ is applied. Indeed, we can treat the values $s_{x,y}$ for $x \in X_C$ and $y \in X(x)$ as probability transitions from the state $x \in X_C$ to the state $y \in X(x)$.

Therefore, for fixed s the solution $q^s = (q^s_{x_{i_1}}, q^s_{x_{i_2}}, \ldots, q^s_{x_{i_n}})$ of the system of linear equations

$$\begin{cases} \sum\limits_{x \in X_C} s_{x,y} q_x + \sum\limits_{z \in X_N} p_{z,y} q_z = q_y, \quad \forall y \in X; \\ \sum\limits_{x \in X_C} q_x + \sum\limits_{z \in X_N} q_z = 1; \end{cases} \tag{2.12}$$

corresponds to the vector of limit probabilities in the ergodic Markov chain determined by the graph $G_s = (X, E_s \cup E_N)$ with the probabilities $p_{x,y}$ for $(x, y) \in E_N$ and $p_{x,y} = s_{x,y}$ for $(x, y) \in E_C$ $(E_C = E \backslash E_N)$. Therefore, for given s the value

$$\psi(s, q^s) = \sum_{x \in X_C} \mu_x q_x + \sum_{z \in X_N} \mu_z q_z$$

expresses the average cost per transition for the dynamical system if the strategy s is applied, i.e.,

$$f_x(s) = \psi(s, q^s), \quad \forall x \in X.$$

So, if we solve the optimization problem (2.10), (2.11) on a perfect network then we find the optimal strategy s^*. □

Remark 2.2 In the case of a perfect network the objective function $\psi(s, q)$ on the feasible set of solutions of the system (2.11) depends only on s, because q_x for $x \in X$ can be uniquely expressed via $s_{x,y}$ ($x \in X_C$, $y \in X$) according to (2.12). Moreover, for perfect networks the condition $q_x \geq 0$ for $x \in X$ in (2.11) holds if $s_{x,y} \geq 0$, $\forall x \in X_C$, $y \in X$. Therefore, the condition $q_x \geq 0$ for $x \in X$ in (2.11) is redundant and can be omitted. This condition is essential only for multichain control problems.

In the following for an arbitrary vertex $y \in X$ we will denote by $X_C^-(y)$ the set of vertices from X_C which contain directed leaving edges $e = (x, y) \in E$ that end in y, i.e., $X_C^-(y) = \{x \in X_C \mid (x, y) \in E\}$; in an analogues way we define the set $X^-(y) = \{x \in X \mid (x, y) \in E\}$.

Based on the lemma above we can prove the following result.

Theorem 2.3 *Let $\alpha^*_{x,y}$ ($x \in X_C$, $y \in X$), q^*_x ($x \in X$) be a basic optimal solution of the following linear programming problem:*
Minimize

$$\overline{\psi}(\alpha, q) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y} \alpha_{x,y} + \sum_{z \in X_N} \mu_z q_z \tag{2.13}$$

subject to

$$\begin{cases} \displaystyle\sum_{x \in X_C^-(y)} \alpha_{x,y} + \sum_{z \in X_N} p_{z,y} q_z = q_y, \quad \forall y \in X; \\ \displaystyle\sum_{x \in X_C} q_x + \sum_{z \in X_N} q_z = 1; \\ \displaystyle\sum_{y \in X(x)} \alpha_{x,y} = q_x, \quad \forall x \in X_C; \\ \alpha_{x,y} \geq 0, \ \forall x \in X_C, \ y \in X; \ \ q_x \geq 0, \ \forall x \in X. \end{cases} \tag{2.14}$$

Then the optimal stationary strategy s^* on a perfect network can be found as follows:

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

where $x \in X_C$, $y \in X(x)$. Moreover, for every starting state $x \in X$ the optimal average cost per transition is equal to $\overline{\psi}(\alpha^*, q^*)$, i.e.,

$$f_x(s^*) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y} \alpha^*_{x,y} + \sum_{z \in X_N} \mu_z q^*_z$$

for every $x \in X$.

Proof To prove the theorem it is sufficient to apply Lemma 2.1 and to show that the bilinear programming problem (2.10), (2.11) with boolean variables $s_{x,y}$ for $x \in X_C$, $y \in X$ can be reduced to the linear programming problem (2.13), (2.14). Indeed, we observe that the restrictions $s_{x,y} \in \{0, 1\}$ in the problems (2.10), (2.11) can be replaced by $s_{x,y} \geq 0$ because the optimal solutions after such a transformation of the problem are not changed. In addition, the restrictions

$$\sum_{y \in X(x)} s_{x,y} = 1, \quad \forall x \in X_C$$

can be changed by the restrictions

$$\sum_{y \in X(x)} s_{x,y} q_x = q_x, \quad \forall x \in X_C$$

because for the perfect network it holds $q_x > 0, \ \forall x \in X_C$.

Based on the properties mentioned above in the problem (2.10), (2.11) we may replace the system (2.11) by the following system

$$\begin{cases} \sum\limits_{x \in X_C^-(y)} s_{x,y} q_x + \sum\limits_{z \in X_N} p_{z,y} q_z = q_y, \quad \forall y \in X; \\ \sum\limits_{x \in X_C} q_x + \sum\limits_{z \in X_N} q_z = 1; \\ \sum\limits_{y \in X(x)} s_{x,y} q_x = q_x, \quad \forall x \in X_C; \\ s_{x,y} \geq 0, \ \forall x \in X_C, \ y \in X; \quad q_x \geq 0, \ \forall x \in X. \end{cases} \tag{2.15}$$

Thus, we may conclude that problem (2.10), (2.11) and problem (2.10), (2.15) 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 (2.10), (2.15) the notations $\alpha_{x,y} = s_{x,y} q_x$ for $x \in X_C$, $y \in X(x)$. This leads to the problem (2.13), (2.14). It is evident that $\alpha_{x,y} \neq 0$ if and only if $s_{x,y} = 1$. Therefore, the optimal stationary strategy s^* can be found according to the rule given in the theorem. □

Remark 2.4 In Theorem 2.3 the linear programming problem (2.13), (2.14) can be changed by the following equivalent linear programming problem:

Minimize

$$\overline{\psi}(\alpha, q) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y}\alpha_{x,y} + \sum_{z \in X_N} \mu_z q_z \tag{2.16}$$

subject to

$$\begin{cases} \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - \sum\limits_{x \in X(y)} \alpha_{y,x} + \sum\limits_{x \in X_N} p_{x,y} q_x = 0, \quad \forall y \in X_C; \\ \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - q_y + \sum\limits_{x \in X_N} p_{x,y} q_x = 0, \quad \forall y \in X_N; \\ \sum\limits_{x \in X_C} \sum\limits_{y \in X(x)} \alpha_{x,y} + \sum\limits_{x \in X_N} q_x = 1; \\ \alpha_{x,y} \geq 0, \ \ \forall x \in X_C, \ \ y \in X; \ \ q_x \geq 0, \ \ \forall x \in X_N. \end{cases} \tag{2.17}$$

This problem is obtained from (2.13), (2.14) if we take into account Remark 2.2 and eliminate q_x for $x \in X_C$ from (2.14). If we solve this problem then we should take into account that $\alpha_{x,y} = s_{x,y} q_x, \ \ \forall x \in X_C, \ \ y \in X(x)$, where $q_x = \sum_{y \in X(x)} \alpha_{x,y}$, $\forall x \in X_C$.

So, if the network (G, X_C, X_N, c, p, x_0) is perfect then we can find the optimal stationary strategy s^* by using the following algorithm.

Algorithm 2.5 Determining the Optimal Stationary Strategy on Perfect Networks

(1) Formulate the linear programming problem (2.13), (2.14) and find a basic optimal solution $\alpha^*_{x,y}$ $(x \in X_C, \ y \in X)$, q^*_x $(x \in X)$.
(2) Fix a stationary strategy s^* where $s^*_{x,y} = 1$ for $x \in X_C, \ y \in X(x)$ if $\alpha^*_{x,y} > 0$; otherwise put $s^*_{x,y} = 0$.

Below an example for determining the optimal control problem on networks by using linear programming is given.

Example Consider a stochastic control problem for which the network is represented in Fig. 2.1, i.e.,

$$G = (X, E), \quad X = \{1, 2, 3, 4\}, \quad X_C = \{1, 2\}, \quad X_N = \{3, 4\},$$
$$E = \{(1, 3), \ (1, 4), \ (2, 3), \ (2, 4), \ (3, 1), \ (3, 4), \ (4, 2), \ (4, 3)\}.$$

The transition cost for directed edges from E and the transition probabilities for directed edges originating in the vertices 3 and 4 are given by:

$$\begin{array}{llll} c_{1,3} = 1, & c_{2,3} = 3, & c_{3,1} = 2, & c_{4,2} = 1, \\ c_{1,4} = 2, & c_{2,4} = 1, & c_{3,4} = 4, & c_{4,3} = 3, \\ p_{3,1} = 0.5, & p_{3,4} = 0.5, & p_{4,2} = 0.5, & p_{4,3} = 0.5\,. \end{array}$$

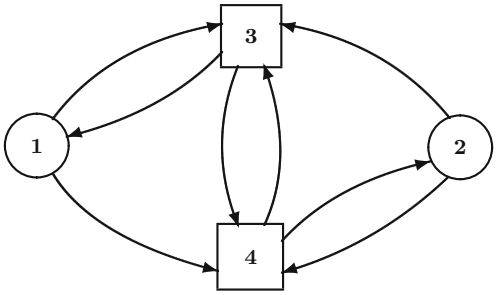

Fig. 2.1 The perfect network for the control problem

We are seeking for the optimal stationary strategy s^* which gives the solution of the problem for an arbitrary starting state $x \in X$.

It is easy to see that the network is perfect and, therefore, we can determine the optimal strategy by solving the linear programming problem (2.13), (2.14).

For this example we have

$$\overline{\psi}(\alpha, q) = c_{1,3}\alpha_{1,3} + c_{1,4}\alpha_{1,4} + c_{2,3}\alpha_{2,3} + c_{2,4}\alpha_{2,4} + \mu_3 q_3 + \mu_4 q_4,$$

where

$$\begin{aligned} \mu_3 &= p_{3,1}c_{3,1} + p_{3,4}c_{3,4} = 0.5 \cdot 2 + 0.5 \cdot 4 = 3, \\ \mu_4 &= p_{4,2}c_{4,2} + p_{4,3}c_{4,3} = 0.5 \cdot 1 + 0.5 \cdot 3 = 2. \end{aligned}$$

So, to determine the optimal stationary strategy s^* we need to solve the linear programming problem:
Minimize

$$\overline{\psi}(\alpha, q) = \alpha_{1,3} + 2\alpha_{1,4} + 3\alpha_{2,3} + \alpha_{2,4} + 3q_3 + 2q_4$$

subject to

$$\begin{cases} 0.5q_3 = q_1, \\ 0.5q_4 = q_2, \\ \alpha_{1,3} + \alpha_{2,3} + 0.5q_4 = q_3, \\ \alpha_{1,4} + \alpha_{2,4} + 0.5q_3 = q_4, \\ \alpha_{1,3} + \alpha_{1,4} = q_1, \\ \alpha_{2,3} + \alpha_{2,4} = q_2, \\ q_1 + q_2 + q_3 + q_4 = 1, \\ q_i \geq 0,\ i = 1, 2, 3, 4; \quad \alpha_{i,j} \geq 0,\ i, j = 1, 2, 3, 4. \end{cases}$$

It is easy to check that the optimal solution of this problem is

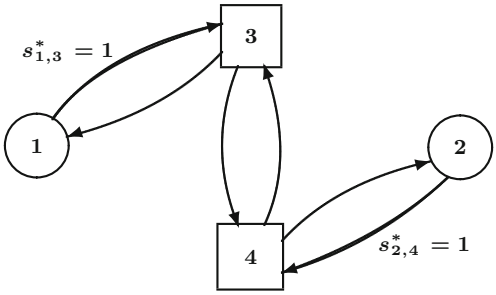

Fig. 2.2 The network induced by the optimal strategy

$$\alpha^*_{1,4} = 0, \quad \alpha^*_{2,3} = 0, \quad \alpha^*_{1,3} = \frac{1}{6}, \quad \alpha^*_{2,4} = \frac{1}{6},$$
$$q^*_1 = \frac{1}{6}, \quad q^*_2 = \frac{1}{6}, \quad q^*_3 = \frac{2}{6}, \quad q^*_4 = \frac{2}{6} \quad \text{and} \quad \varphi(\alpha^*, q^*) = 2.$$

So, $s^*_{1,4} = 0,\ s^*_{2,3} = 0,\ s^*_{1,3} = 1,\ s^*_{2,4} = 1$.

In Fig. 2.2 a network is presented which corresponds to an optimal stationary strategy $s^*_{1,3} = 1, s^*_{2,4} = 1$.

2.2.3 Remark on the Application of the Unichain Linear Programming Model for an Arbitrary Network

The linear programming problem (2.13), (2.14) can be solved on an arbitrary decision network (G, X_C, X_N, c, p). A basic optimal solution α^*, q^* determines the strategy

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

and a subset $X^* = \{x \in X \mid q_{x^*} > 0\}$, where s^* provides the optimal average cost per transition for the dynamical system $\mathbb{L}$ when it starts transitions in the states $x_0 \in X^*$.

This means that for an arbitrary network Algorithm 2.5 determines the optimal stationary strategy of the problem only in the case if the system starts transitions in the states $x \in X^*$. So, in the general case the algorithm finds a strategy s^* and a distinct positive recurrent class X^* in X with the minimal average cost per transition of the system $\mathbb{L}$ for an arbitrary starting state $x_0 \in X^*$.

For a unichain control problem Algorithm 2.5 determines the strategy s^* and the recurrent class X^*. In this case the remaining states $x \in X \backslash X^*$ in X correspond to transient states and the optimal stationary strategies in the states $x \in X \backslash X^*$ can be chosen in order to reach X^*. Therefore, the linear programming model (2.13), (2.14) can be used for determining the optimal stationary strategy for an arbitrary unichain control problem.

2.2.4 Determining the Solutions for an Arbitrary Unichain Control Problem and for the Deterministic Case

As we have noted the linear programming model (2.13), (2.14) can be used for studying the control problem on a network of arbitrary structure. Here we show how to use the linear programming model (2.13), (2.14) for determining the optimal stationary strategies of the control problem in the following two cases:

(1) the network is not perfect but for an arbitrary stationary strategy s the matrix P^s corresponds to a recurrent Markov chain;
(2) the network contains only controllable states, i.e., $X_N = \emptyset$.

First let us analyze the problem in the case (1). In this case an arbitrary strategy s in G generates a graph G_s with unique strongly connected components $G'_s = (X'_s, E'_s)$ that can be reached from any vertex $x \in X$. The optimal stationary strategy s^* in G can be found from a basic optimal solution by fixing $s^*_{x,y} = 1$ for the basic variables. This means that in G we can find the optimal stationary strategy as follows:

We solve the linear programming problem (2.13), (2.14) and find a basic optimal solution α^*, q^*. Then we find the subset of vertices $X^* = \{x \in X \mid q^*_x > 0\}$ which in G corresponds to a strongly connected subgraph $G^* = (X^*, E^*)$. On this subgraph we determine the optimal solution of the problem using the algorithm described in the previous section. It is evident that if $x_0 \in X^*$ then we obtain the solution of the problem with fixed starting state x_0. To determine the solution of the problem for an arbitrary starting state we may select successively vertices $x \in X \backslash X^*$ which contain outgoing directed edges that end in X^* and will add them at each time to X^* using the following rule:

- if $x \in X_C \cap (X \backslash X^*)$ then we fix an directed edge $e = (x, y)$, put $s^*_{x,y} = 1$ and change X^* by $X^* \cup \{x\}$;
- if $x \in X_N \cap (X \backslash X^*)$ then change X^* by $X^* \cup \{x\}$.

Thus, in the case (1) we can determine the optimal stationary strategy of the control problem on the network (G, X_C, X_N, c, p, x_0).

In the case (2) ($X_N = \emptyset$) we have a deterministic model and the linear programming problem (2.13), (2.14) becomes the linear programming problem from [65, 117]. Thus, the linear programming model generalizes the deterministic model from [65, 117] and from Theorem 2.3 we obtain the following result.

Lemma 2.6 *Let $G = (X, E)$ be a strongly connected directed graph with $X_N = \emptyset$ and let $\alpha^*_{x,y}$, $(x, y) \in E$ be the basic optimal solution of the linear programming problem:*
Minimize

$$\overline{\psi}(\alpha) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y} \alpha_{x,y} \tag{2.18}$$

subject to

$$\begin{cases} \sum\limits_{x\in X^-(y)} \alpha_{x,y} - \sum\limits_{z\in X(y)} \alpha_{y,z} = 0, \quad \forall y \in X; \\ \sum\limits_{x\in X} \sum\limits_{y\in X(x)} \alpha_{x,y} = 1; \\ \alpha_{x,y} \geq 0, \quad \forall (x, y) \in E. \end{cases} \tag{2.19}$$

Then the subgraph $G' = (X', E')$ generated by the directed edges $(x, y) \in E$ with $\alpha^*_{x,y} > 0$ has a structure of a directed cycle and an optimal stationary strategy s^* for the control problem on G with a given starting state x_0 can be found as follows:

- *fix a simple directed path which connects x_0 with the directed cycle G' and find the set of edges E'' of this directed path;*
- *fix the stationary strategy s^* where $s^*_{x,y} = 1$ if $(x, y) \in E' \cup E''$; otherwise put $s^*_{x,y} = 0$.*

Proof If $X_N = \emptyset$ then problem (2.13), (2.14) is transformed into the following problem:
Minimize (2.18) subject to

$$\begin{cases} \sum\limits_{x\in X^-(y)} \alpha_{x,y} = q_y, \quad \forall y \in X; \\ \sum\limits_{x\in X} q_y + \sum\limits_{z\in X_N} q_z = 1; \\ \sum\limits_{y\in X(x)} \alpha_{x,y} = q_x, \quad \forall x \in X; \\ \alpha_{x,y} \geq 0, \quad \forall x, y \in X; \quad q_x \geq 0, \quad \forall x \in X. \end{cases} \tag{2.20}$$

After the elimination of q_x and q_y from the system (2.20) we obtain the system (2.19). In such a way we obtain that (2.18), (2.19) becomes the mean cost cycle problem on G and the algorithm from the lemma above determines the optimal solution of the problem. □

Based on the lemma above we can propose the following algorithm for finding the solution of the problem in the case $X_N = \emptyset$.

Algorithm 2.7 Determining the Optimal Solution for the Deterministic Control Problem

1. Formulate the linear programming problem (2.18), (2.19) and find a basic optimal solution $\alpha^*_{x,y}$ and the corresponding directed graph $G' = (X', E')$ which has the structure of a directed cycle;
2. Fix a simple directed path which connects x_0 with the directed cycle G' and find the set of edges E'' of this directed path;

3. Fix a stationary strategy s^* where $s^*_{x,y} = 1$ if $(x, y) \in E' \cup E''$; otherwise put $s^*_{x,y} = 0$.

So, the deterministic control problem can be efficiently solved on an arbitrary network if $X_N = \emptyset$.

2.2.5 Dual Linear Programming for the Unichain Control Problem and an Algorithm for Determining the Optimal Strategies

For the linear programming model (2.16), (2.17) we consider the following dual problem:
Maximize

$$\overline{\psi}'(\varepsilon, \omega) = \omega \tag{2.21}$$

subject to

$$\begin{cases} \varepsilon_x - \varepsilon_y + \omega \le c_{x,y}, \quad \forall\, x \in X_C,\ y \in X(x); \\ \varepsilon_x - \sum\limits_{z \in X} p_{x,z}\varepsilon_z + \omega \le \mu_x, \quad \forall\, x \in X_N. \end{cases} \tag{2.22}$$

Remark 2.8 The conditions $q_y \ge 0$, $\forall x \in X_N$ in the unichain primal linear programming problem are redundant. Therefore, the constraints 2.22 in the problem (2.21), (2.22) can be replaced by the following constraints

$$\begin{cases} \varepsilon_x - \varepsilon_y + \omega \le c_{x,y}, \quad \forall\, x \in X_C,\ y \in X(x); \\ \varepsilon_x - \sum\limits_{z \in X} p_{x,z}\varepsilon_z + \omega = \mu_x, \quad \forall\, x \in X_N. \end{cases} \tag{2.23}$$

The optimal stationary strategies of the unichain control problem correspond to basic optimal solutions of this problem and can be found by using the following theorem.

Theorem 2.9 *An arbitrary optimal solution ε^*_x $(x \in X)$, ω^* of the problem (2.21), (2.22) for a unichain control model on the network (G, X_C, X_N, c, p) possesses the following property:*

(1) $\min\limits_{y \in X_(x)} \{c_{x,y} + \varepsilon^*_y - \varepsilon^*_x - \omega^*\} = 0, \quad \forall x \in X_C;$

(2) $\mu_x + \sum\limits_{z \in X(x)} p_{x,z}\varepsilon^*_z - \varepsilon^*_x - \omega^* = 0, \quad \forall x \in X_N;$

(3) a stationary strategy $s^* : X_C \to X$ is optimal if and only if $(x, s^*(x)) \in E^*_C$, $\forall x \in X_C$, where

$$E_C^* = \{e = (x, y) \in E_C \mid c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - \omega^* = 0\}.$$

The value ω^* is equal to the optimal average cost in the unichain control problem on the network (G, X_C, X_N, c, p).

Proof The properties (1) and (2) of the theorem represent the optimality conditions for the dual linear programming problem (2.21), (2.22). If $\alpha_{x,y}^*$, $(x \in X_C, y \in X(x))$, q_x^* $(x \in X)$ is a basic solution of the primal problem (2.16), (2.17), where $\alpha_{x,y}^* = s_{x,y}^* q_x^*$, $q^* = \sum_{y \in X(x)} \alpha_{x,y}$, then we can take $s_{x,y}^* = 1$ for $(x, y) \in E_C$ that satisfies the conditions (1), (2) and $s_{x,y} = 0$ in the other case. This means that an optimal stationary strategy in G is determined by the map $s^* : X_C \to X$ for which $(x, s^*(x)) \in E_C^*$, $\forall x \in X_C$. □

Corollary 2.10 *Each subset $E_{s^*} = \{e = (x, s^*(x)) \in E_C^* \mid x \in X_C\}$ in G generates a subgraph $G_{s^*} = (X, E_{s^*} \cup E_C)$ that corresponds to a Markov unichain, i.e., G_{s^*} contains a unique strongly connected component that is reachable from every $x \in X$. The values of the boolean variable $s_{x,y}^*$, $x \in X$, $y \in X(x)$ that correspond to an optimal solution of the problem can be found by fixing*

$$s_{x,y}^* = \begin{cases} 1, & \text{if } (x, y) \in E_{s^*}; \\ 0, & \text{if } (x, y) \notin E_{s^*}. \end{cases}$$

Corollary 2.11 *Let s be an arbitrary strategy for the control problem on the network (G, X_C, X_N, c, p) and $P^s = (p_{x,y}^s)$ be the transition probability matrix induced by this strategy,*

$$p_{x,y}^s = \begin{cases} p_{x,y}, & \text{if } x \in X_N \text{ and } y \in X; \\ 1, & \text{if } x \in X_C \text{ and } y = s(x); \\ 0, & \text{if } x \in X_C \text{ and } y \neq s(x). \end{cases}$$

Then in the Markov process induced by this transition probability matrix it holds

$$q_x^s \left(\mu_x^s + \varepsilon_x^s - \sum_{z \in X} p_{x,z}^s \varepsilon_z^s - \omega^s \right) = 0 \quad \forall x \in X,$$

where q_x^s is a limiting probability in the state $x \in X$ and $\mu_x^s = \sum_{y \in X(x)} p_{x,y}^s c_{x,y}$.

From Theorem 2.9 we can make the following conclusions. For an arbitrary unichain control problem there exist a function $\varepsilon^* : X \to \mathbb{R}$ and a value ω^* that satisfy the conditions

(1) $\overline{c}_{x,y} = c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - \omega^* \geq 0, \quad \forall x \in X_C, \ \forall y \in X(x);$

(2) $\min\limits_{y \in X} \overline{c}_{x,y} = 0, \quad \forall x \in X_C;$

(3) $\overline{\mu}_x = \mu_x + \sum\limits_{y \in X} p_{x,y} \varepsilon_x^* - \varepsilon_x^* - \omega_x^* = 0, \quad \forall x \in X_N.$

If in the decision network (G, X_C, X_N, c, p) we change the cost function c by $\overline{c}$ then we obtain a new control problem on the network $(G, X_C, X_N, \overline{c}, p)$. Such a transformation of the cost function in the control problem does not change the optimal stationary strategies. In the new control problem the cost function $\overline{c}$ satisfies the conditions $\min_{y \in X(x)} \overline{c}_{x,y} = 0, \ \forall x \in X_C$ and $\overline{\mu}_x = 0, \ \forall x \in X_N$. For this problem the optimal average cost $\overline{\omega}_x^*$ for every $x \in X$ is equal to zero and an optimal stationary strategy can be found by fixing an arbitrary map s^* such that $(x, s^*(x)) \in E_C^*$, where $E_C^* = \{(x, y) \in E_C \mid \overline{c}_{x,y} = 0\}$.

We call the cost function $\overline{c}_{x,y} = c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - \omega^*, \ (x, y) \in E$ a *potential transformation* induced by the *potential function* $\varepsilon^* : X \to \mathbb{R}$ and the values ω_x^* for $x \in X$. Furthermore, we call the new problem with the cost function $\overline{c}$ a *control problem in canonical form*.

2.2.6 The Potential Transformation and Optimality Conditions for Multichain Control Problems

The aim of this section is to formulate and prove the optimality conditions for an average multichain stochastic control problem. For this reason we extend the notions of the potential function and *potential transformation* for a multichain control problem and study their main properties. Based on these properties we prove the optimality conditions and show how to reduce the average multichain control problem to an auxiliary one in canonical form for which the optimal solutions can easily be found. We show that such a transformation of the control problem into an auxiliary problem in canonical form always exists. Finally, we show that the problem of determining optimal stationary strategies in a multichain control problem can be formulated as a linear programming problem.

We define the *decision network in canonical form* $(G, X_C, X_N, \overline{c}, p)$ for a multichain control problem on the network (G, X_C, X_N, c, p) by using the potential transformation

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon_y - \varepsilon_x - h_x, \quad \forall x \in X, \ \forall y \in X(x), \tag{2.24}$$

where the function $\varepsilon : X \to \mathbb{R}$ and the values h_x for $x \in X$ satisfy the conditions:

(1) $\overline{c}_{x,y} = c_{x,y} + \varepsilon_y - \varepsilon_x - h_x \geq 0, \quad \forall x \in X_C, \ y \in X(x)$;

(2) $\min\limits_{y \in X} \overline{c}_{x,y} = 0, \quad \forall x \in X_C$;

(3) $\overline{\mu}_x = \mu_x + \sum\limits_{y \in X} p_{x,y}\varepsilon_y - \varepsilon_x - h_x = 0, \quad \forall x \in X_N$;

(4) $h_x = \min\limits_{y \in X(x)} h_y, \quad \forall x \in X_C, \ \forall y \in X(x)$;

(5) $h_x = \sum_{y \in X} p_{x,y} h_y, \quad \forall x \in X_N$

(6) $E_h(x) \cap E_{\overline{c}}(x) \neq \emptyset$, where

$$E_h(x) = \left\{ (x, y) \in E_C \mid y \in \operatorname*{argmin}_{z \in X(x)} \{h_z\} \right\}, \quad x \in X_C$$

and

$$E_{\overline{c}}(x) = \left\{ (x, y) \in E_C \mid y \in \operatorname*{argmin}_{z \in X(x)} \{\overline{c}_{x,z}\} \right\}, \quad x \in X_C.$$

In general, the potential transformation (2.24) can also be considered for an arbitrary network. However, the optimal stationary strategies in the control problem after such a potential transformation may differ from the optimal stationary strategies in the initial network. The potential transformation with the properties mentioned above preserves the optimal strategy of the multichain control problem.

If the decision network in canonical form is known then the optimal stationary strategy for the stochastic multichain control problem can be found in a similar way as for the unichain case of the problem, i.e., we fix a strategy $s^* : X_C \to X$ such that $(x, s^*(x)) \in E^*_{\overline{c}}$. Moreover, the potential transformation $\overline{c}$ that satisfies the conditions (1)–(6) gives the values of the optimal average costs $\omega^*_x = h_x$ in the states $x \in X$ for a multichain control problem on the network (G, X_C, X_N, c, p).

In the following we show that for an arbitrary network (G, X_C, X_N, c, p) that there exists a network in canonical form $(G, X_C, X_N, \overline{c}, p)$ that obtains the optimal stationary strategy s^* and the optimal average costs ω^*_x for $x \in X$. We ground all these results on the basis of the following optimality principle for a multichain control problem.

Theorem 2.12 *For an arbitrary decision network* (G, X_C, X_N, c, p) *there exists a potential transformation*

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon^*_y - \varepsilon^*_x - h^*_x, \quad \forall x \in X, \ y \in X(x)$$

of the cost function c that satisfies the following conditions:

(1) $\overline{c}_{x,y} = c_{x,y} + \varepsilon^*_y - \varepsilon^*_x - h^*_x \geq 0, \quad \forall x \in X_C, \ y \in X(x);$

(2) $\min_{y \in X} \overline{c}_{x,y} = 0, \quad \forall x \in X_C;$

(3) $\overline{\mu}_x = \mu_x + \sum_{y \in X} p_{x,y} \varepsilon^*_y - \varepsilon^*_x - h^*_x = 0, \quad \forall x \in X_N;$

(4) $h^*_x = \min_{y \in X(x)} h^*_y, \quad \forall x \in X_C, \ \forall y \in X(x);$

(5) $h_x^* = \sum_{y \in X} p_{x,y} h_y^*, \quad \forall x \in X_N;$

(6) $E_{h^*}^*(x) \cap E_{\overline{c}}^*(x) \neq \emptyset, \ \forall x \in X_C,$ *where*

$$E_{h^*}^*(x) = \left\{ (x, y) \in E_C \mid \ y \in \operatorname*{argmin}_{z \in X(x)} \{h_z\} \right\}, \ x \in X_C$$

and

$$E_{\overline{c}}^*(x) = \left\{ (x, y) \in E_C \mid \ y \in \operatorname*{argmin}_{z \in X(x)} \{\overline{c}_{x,z}\} \right\}, \ x \in X_C.$$

The values ε_x^* *for* $x \in X$ *correspond to a basic solution of the system of linear equations*

$$\begin{cases} c_{x,y} + \varepsilon_y - \varepsilon_x - h_x^* = 0, \ \ \forall x \in X_C, \ (x, y) \in E_{h^*}^*(x); \\ \mu_x + \sum_{y \in X} p_{x,y} \varepsilon_y - \varepsilon_x - h_x^* = 0, \quad \forall x \in X_N \end{cases} \tag{2.25}$$

and determines the decision network in canonical form $(G, X_C, X_N, \overline{c}, p)$ *for the control problem on the network* (G, X_C, X_N, c, p)*, where* $\overline{c}_{x,y} = c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - h_x^*$, $\forall x \in X, \ y \in X(x)$.

The values h_x^* *for* $x \in X$ *coincide with the corresponding optimal average costs* ω_x^* *for* $x \in X$ *and an optimal stationary strategy for the control problem on the network can be found by fixing an arbitrary map* $s^* : X_C \to X$ *such that* $(x, s^*(x)) \in E_{h^*}^*(x) \cap E_{\overline{c}}^*(x), \ \forall x \in X_C.$

This theorem is tightly connected with the existence of the solutions for the *bias equations in average Markov decision processes* (see [115, 140]). In the terms of bias equations this theorem can be formulated in the following way:

Theorem 2.13 *The system of equations*

$$\begin{cases} \varepsilon_x + h_x = \min_{y \in X} \{c_{x,y} + \varepsilon_y\}, \quad \forall x \in X_C; \\ \varepsilon_x + h_x = \mu_x + \sum_{y \in X} p_{x,y} \varepsilon_y, \ \ \forall x \in X_N \end{cases} \tag{2.26}$$

has solutions with respect to ε_x *for* $x \in X$ *under the set of solutions of the following system of equations*

$$\begin{cases} h_x = \min_{y \in X(x)} h_y, \qquad \forall x \in X_C; \\ h_x = \sum_{y \in X(x)} p_{x,y} h_x, \quad \forall x \in X_N. \end{cases} \tag{2.27}$$

If ε_x^*, h_x^* $(x \in X)$ is the solution of these equations then h_x^* for $x \in X$ coincides with the optimal average costs ω_x^*.

To prove Theorem 2.12 we need some auxiliary results.

Let $s : X \to X$ be a feasible strategy for the control problem on the decision network (G, X_C, X_N, c, p) and $P^s = (p^s_{x,y})$ be the transition probability matrix of the Markov chain induced by the strategy s, i.e.,

$$p^s_{x,y} = \begin{cases} p_{x,y}, & \text{if } x \in X_N \text{ and } y = X(x); \\ 1, & \text{if } x \in X_C \text{ and } y = s(x); \\ 0, & \text{if } x \in X_C \text{ and } y \neq s(x). \end{cases} \tag{2.28}$$

Denote by $Q^s = (q^s_{x,y})$ the limit matrix in the Markov chain with probability transition matrix P^s and by $X^s_1, X^s_2, \ldots, X^s_k$ the corresponding irreducible sets in this Markov chain.

Lemma 2.14 *Let ω^s_x be the average cost per transition of the system for a feasible strategy $s : X_C \to X$ of the control problem on the decision network (G, X_C, X_N, c, p). Then for an arbitrary potential function $\varepsilon : X \to \mathbb{R}$ and arbitrary real values h_x for $x \in X$ the average cost per transition $\overline{\omega}^s_x$ of the system on the potential transformed network $(G, X_C, X_N, p, \overline{c})$ satisfies the condition*

$$\overline{\omega}^s_x = \omega^s_x - \sum_{z \in X} q^s_{x,z} h_z, \quad \forall x \in X. \tag{2.29}$$

Proof Let s be a feasible stationary strategy of the control problem. Consider a potential transformation $\overline{c}_{x,y} = c_{x,y} + \varepsilon_y - \varepsilon_x - h_x, (x, y) \in E$ determined by an arbitrary function $\varepsilon : X \to \mathbb{R}$ and arbitrary real values h_z for $z \in X$. Then after the potential transformation the average cost $\overline{\omega}^s_x$ for an arbitrary $x \in X$ can be calculated as follows:

$$\begin{aligned} \overline{\omega}^s_x &= \sum_{z \in X} \overline{\mu}^s_z q^s_{x,z} = \sum_{z \in X} \sum_{y \in X(z)} p^s_{z,y} \overline{c}_{z,y} q^s_{x,z} \\ &= \sum_{z \in X} \sum_{y \in X(z)} p^s_{z,y} (c_{z,y} + \varepsilon^s_y - \varepsilon^s_z - h_z) q^s_{x,z} = \sum_{z \in X} \sum_{y \in X(z)} p^s_{z,y} c_{z,y} q^s_{x,z} \\ &\quad + \sum_{z \in X} q^s_{x,z} \sum_{y \in X(z)} p^s_{z,y} \varepsilon^s_y - \sum_{z \in X} q^s_{x,z} \sum_{y \in X(z)} p^s_{z,y} \varepsilon^s_z - \sum_{z \in X} q^s_{x,z} h_z \sum_{y \in X(z)} p^s_{z,y} \\ &= \omega^s_x + \sum_{z \in X} q^s_{x,z} \left(\sum_{y \in X(z)} p^s_{z,y} \varepsilon_y - \sum_{y \in X(z)} p^s_{z,y} \varepsilon_z \right) - \sum_{z \in X} q^s_{x,z} h_z, \end{aligned}$$

i.e., we have

$$\overline{\omega}^s_x = \omega^s_x + \sum_{z \in X} q^s_{x,z} \left(\sum_{y \in X(x)} p^s_{z,y} \varepsilon^s_y - \varepsilon^s_z \right) - \sum_{z \in X} q^s_{x,z} h_z, \quad \forall x \in X. \tag{2.30}$$

Now we show that for an arbitrary strategy s it holds

$$\sum_{z\in X} q^s_{x,z}\Big(\sum_{y\in X(z)} p^s_{z,y}\varepsilon^s_y - \varepsilon^s_z\Big) = 0, \quad \forall x \in X. \tag{2.31}$$

Let $X^s_1, X^s_2, \ldots, X^s_k$ be the corresponding irreducible sets in the Markov chain induced by the strategy s. Then in the graph $G_s = (X, E_s \cup E_N)$ each subset X^s_i of X generates a strongly connected graph that corresponds to a distinct irreducible Markov chain and in each irreducible set the average costs for an arbitrary starting state is the same.

If we denote by $\omega^{s,i}$ the average cost for the corresponding states in the irreducible sets X^s_i then we have

$$\begin{aligned}
&\sum_{z\in X} q^s_{x,z}\Big(\sum_{y\in X(z)} p^s_{z,y}\varepsilon^s_y - \varepsilon^s_z\Big)\\
&= \sum_{z\in X} q^s_{x,z}\Big(\Big(\mu^s_z + \sum_{y\in X(z)} p^s_{z,y}\varepsilon^s_y - \varepsilon^s_z - \omega^s_z\Big) + (\omega^s_z - \mu^s_z)\Big)\\
&= \sum_{i=1}^{k}\sum_{z\in X^s_i} q^s_{x,z}\Big(\Big(\mu^s_z + \sum_{y\in X(z)} p^s_{z,y}\varepsilon^s_y - \varepsilon^s_z - \omega^{s,i}\Big) + (\omega^{s,i} - \mu^s_z)\Big)\\
&= \sum_{i=1}^{k}\sum_{z\in X^s_i} q^s_{x,z}\Big(\mu^s_z + \sum_{y\in X(z)} p^s_{z,y}\varepsilon^s_y - \varepsilon^s_z - \omega^{s,i}\Big) + \sum_{i=1}^{k}\sum_{z\in X^s_i} q^s_{x,z}(\omega^{s,i} - \mu^s_z).
\end{aligned}$$

Here, according to Corollary 2.11 it holds

$$\mu^s_z + \sum_{y\in X(z)} p^s_{z,y}\varepsilon^s_y - \varepsilon^s_z - \omega^{s,i} = 0, \quad \forall z \in X^s_i,\ i = 1, 2, \ldots, k.$$

Therefore, we obtain

$$\begin{aligned}
\sum_{z\in X} q^s_{x,z}\Big(\sum_{y\in X(z)} p^s_{z,y}\varepsilon^s_y - \varepsilon^s_z\Big) &= \sum_{i\in 1}^{k}\sum_{z\in X^s_i} q^s_{x,z}(\omega^{s,i} - \mu^s_z)\\
&= \sum_{i=1}^{k}\omega^{s,i}\sum_{z\in X^s_i} q^s_{x,z} - \sum_{i=1}^{k}\sum_{z\in X^s_i} q^s_{x,z}\mu^s_z = \sum_{i=1}^{k}\omega^{s,i} - \sum_{i=1}^{k}\omega^{s,i} = 0.
\end{aligned}$$

So, condition (2.31) holds.
If we introduce (2.31) in (2.30) then we obtain (2.29). □

Corollary 2.15 *Let s be an arbitrary feasible strategy for the control problem on the network (G, X_C, X_N, c, p) and $Q^s = (q^s_{x,y})$ be the matrix of limiting probabilities in the Markov chain induced by the strategy s. Then for an arbitrary potential function $\varepsilon : X \to \mathbb{R}$ the following condition holds*

$$\sum_{z \in X} q^s_{x,z} \left(\sum_{y \in X(x)} p^s_{z,y} \varepsilon^s_y - \varepsilon^s_z \right) = 0, \quad \forall x \in X. \tag{2.32}$$

If in (2.24) we fix $h_x = h, \ \forall x \in X$ then we obtain the following potential transformation

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon_y - \varepsilon_x - h, \quad \forall x \in X, \ \forall y \in X(x), \tag{2.33}$$

In this case from Lemma 2.14 we obtain the following result.

Corollary 2.16 *Let ω^s_x be the average cost per transition of the system for a feasible strategy $s : X_C \to X$ of the control problem on the decision network (G, X_C, X_N, c, p). Then for an arbitrary potential function $\varepsilon : X \to \mathbb{R}$ and $h \in \mathbb{R}$ the average cost per transition $\overline{\omega}^s_x$ of the system on the potential transformed network $(G, X_C, X_N, \overline{c}, p)$ satisfies the condition*

$$\overline{\omega}^s_x = \omega^s_x - h, \quad \forall x \in X, \ \forall s. \tag{2.34}$$

Corollary 2.16 shows that an arbitrary control problem with average cost criterion can be transformed into a similar one where the transition cost function $\overline{c}$ is nonnegative or positive. Indeed, if we take an arbitrary function $\varepsilon : X \to \mathbb{R}$ and $h = -M$, where $M \geq \max_{(x,y) \in E} |c_{x,y}|$, then the cost function $\overline{c}$ in the control problem becomes nonnegative or positive.

Lemma 2.17 *Assume that for a fixed strategy s the values $h^s_x, \ x \in X$ satisfy the condition*

$$h^s_x - \sum_{y \in X(x)} p^s_{x,y} h^s_y = 0, \quad \forall x \in X. \tag{2.35}$$

Then for an arbitrary potential function $\varepsilon : X \to \mathbb{R}$ the average cost $\overline{\omega}^s_x$ in the control problem on the network $(C, X_C, X_N, \overline{c}, p)$ with a transformed potential cost function

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon_y - \varepsilon_x - h^s_x, \quad \forall x \in X, \ \forall y \in X(x)$$

can be calculated using the following formula

$$\overline{\omega}^s_x = \omega^s_x - h^s_x, \quad \forall x \in X. \tag{2.36}$$

If h_x^s for $x \in X$ satisfies the condition

$$h_x^s - \sum_{y \in X(x)} p_{x,y}^s h_y^s \leq 0, \quad \forall x \in X, \tag{2.37}$$

then

$$\overline{\omega}_x^s \geq \omega_x^s - h_x^s, \quad \forall x \in X. \tag{2.38}$$

If h_x^s for $x \in X$ satisfies the condition

$$h_x^s - \sum_{y \in X(x)} p_{x,y}^s h_y^s \geq 0, \quad \forall x \in X, \tag{2.39}$$

then

$$\overline{\omega}_x^s \leq \omega_x^s - h_x, \quad \forall x \in X, \ \forall s \in \mathbb{S}. \tag{2.40}$$

Proof According to Lemma 1.25 (see Eqs. (1.55), (1.56)) the condition (2.35) implies

$$h_x^s = \sum_{y \in X(x)} q_{x,y}^s h_y^s, \quad \forall x \in X. \tag{2.41}$$

If we introduce (2.41) in (2.29) then we obtain (2.35). In the case if h_x^s for $x \in X$ satisfies (2.37) we obtain $h_x^s \leq \sum_{y \in X(x)} p_{x,y}^s h_y^s$. This implies (2.38). If h_x^s for $x \in X$ satisfies (2.39) then we obtain $h_x^s \geq \sum_{y \in X(x)} p_{x,y}^s h_y^s$. This implies (2.40). □

Lemma 2.18 *Let s be an arbitrary stationary strategy for the control problem on the network (G, X_C, X_N, c, p) and $P^s = (p_{x,y}^s)$ be the probability transition matrix induced by the strategy s, i.e., the elements $p_{x,y}^s$ of this matrix are defined according to (2.28). Then the system of linear equations*

$$\begin{cases} \mu_x^s + \sum_{y \in X} p_{x,y}^s \varepsilon_y^s - \varepsilon_x^s - h_x^s = 0, & \forall x \in X; \\ h_x^s - \sum_{y \in X(x)} p_{x,y}^s h_y^s = 0, & \forall x \in X; \end{cases} \tag{2.42}$$

has solutions. Moreover, if

$$h_x^s - \sum_{y \in X(x)} p_{x,y}^s h_y^s \leq 0, \quad \forall x \in X \tag{2.43}$$

then

$$\mu_x^s + \sum_{y \in X} p_{x,y} \varepsilon_y^s - \varepsilon_x^s - h_x^s \geq 0, \quad \forall x \in X; \tag{2.44}$$

if

$$h_x^s - \sum_{y \in X(x)} p_{x,y}^s h_y^s \geq 0, \quad \forall x \in X \tag{2.45}$$

then

$$\mu_x^s + \sum_{y \in X} p_{x,y} \varepsilon_y^s - \varepsilon_x^s - h_x^s \leq 0, \quad \forall x \in X. \tag{2.46}$$

Proof We shall use the vector representation of the system (2.42). Denote by μ, h and ε the vectors with the corresponding components μ_x, h_x and ε_x for $x \in X$. Additionally, assume that the matrix P^s is represented in canonical form as it is defined in Sect. 1.1.3, i.e.,

$$P^s = \begin{pmatrix} P_1^s & 0 & \dots & 0 & 0 \\ 0 & P_2^s & \dots & 0 & 0 \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ 0 & 0 & \dots & P_k^s & 0 \\ W_1^s & W_2^s & \dots & W_k^s & W_{k+1}^s \end{pmatrix},$$

where P_r^s, $r = 1, 2, \dots, k$ represent the submatrices of P^s that corresponds to the ergodic classes X_r^s of the Markov multichain and W_r^s represent the submatrices of P^s that give the probability transitions from the states $x \in X \setminus (\bigcup_{r=1}^k X_r^s)$ to the states X_r^s; the elements of the matrix W_{k+1}^s represent the probability transitions $p_{x,y}$ between the states $x, y \in \bigcup_{r=1}^k X_r^s$. For each class X_r^s we shall use the vectors $\mu^{s,r}$, $h^{s,r}$ and $\varepsilon^{s,r}$ with the corresponding components $\mu_x^{s,r}$, $h_x^{s,r}$ and $\varepsilon_x^{s,r}$ for $x \in X_r^s$. Using these notations we can write the system (2.42) as follows

$$\begin{cases} \mu^{s,r} - (I^r - P_r^s)\varepsilon^{s,r} - h^{s,r} = 0, \quad r = 1, 2, \dots, k; \\ \mu^{s,k+1} - \sum_{r=1}^k (I^{k+1} - W_r^s)\varepsilon^{s,r} + (I^{k+1} - W_{k+1}^s)\varepsilon^{s,k+1} - h^{s,k+1} = 0; \\ (I^r - P_r^s)h^{s,r} = 0, \quad r = 1, 2, \dots, k; \\ \sum_{r=1}^k (I^r - W_r^s)h^{s,r} + (I^{k+1} - W_{k+1}^s)h^{s,r} = 0. \end{cases} \tag{2.47}$$

In this system each equation

$$\mu^{s,r} - (I^r - P_r^s)\varepsilon^{s,r} - h^{s,r} = 0$$

that corresponds to the class X_r^s, $r \in \{1, 2, \dots, k\}$ has a solution. This solution can be found on the bases of Theorem 2.9. According to this theorem we obtain

$h_x^{s,r} = \omega^{s,r}$, $\forall x \in X_r^s$, where $\omega^{s,r}$ is the average cost of the ergodic class X_r^s. In (2.47) each equation

$$(I^r - P_r^s)h^{s,r} = 0, \quad r \in \{1, 2, \ldots, k\}$$

is redundant and therefore can be deleted. Thus, from the last equation of (2.47) we can determine

$$h^{s,r} = -(I^{k+1} - W_{k+1})^{-1} \sum_{r=1}^{k} (I^r - W_r^s)h^{s,r}.$$

Note that for $(I^{k+1} - W_{k+1})$ there always exists the inverse matrix (see [7, 21, 115]). If we introduce this expression in the equation

$$\mu^{s,k+1} - \sum_{r=1}^{k} (I^{k+1} - W_r^s)\varepsilon^{s,r} + (I^{k+1} - W_{k+1}^s)\varepsilon^{s,k+1} - h^{s,k+1} = 0$$

of the system (2.47) then we can determine uniquely $\varepsilon^{s,k+1}$. So, the system (2.42) obtains solutions.

The second part of the lemma follows from the procedure given above to determine the solution of the system (2.42).

The condition (2.43) implies

$$\mu^{s,k+1} - \sum_{r=1}^{k} (I^{k+1} - W_r^s)\varepsilon^{s,r} + (I^{k+1} - W_{k+1}^s)\varepsilon^{s,k+1} - h^{s,k+1} \geq 0$$

and the condition (2.45) implies

$$\mu^{s,k+1} - \sum_{r=1}^{k} (I^{k+1} - W_r^s)\varepsilon^{s,r} + (I^{k+1} - W_{k+1}^s)\varepsilon^{s,k+1} - h^{s,k+1} \leq 0.$$

In (2.42) the solution of the system of equations

$$\mu^{s,r} - (I^r - P_r^s)\varepsilon^{s,r} - h^{s,r} = 0, \quad r = 1, 2, \ldots, k$$

does not depend on the conditions $(I^r - P_r^s)h^{s,r} \leq 0$ and $(I^r - P_r^s)h^{s,r} \geq 0$. So, the lemma holds. □

Corollary 2.19 *For an arbitrary stationary strategy s on the decision network (G, X_C, X_N, c, p) there exist ε_x^s and h_x^s for $x \in X$ that satisfy the conditions*

(1) $c_{x,y} + \varepsilon_y^s - \varepsilon_x^s - h_x^s = 0, \quad \forall x \in X_C, y = s(x);$

(2) $\mu_x + \sum_{y \in X(x)} p_{x,y}\varepsilon_y^s - \varepsilon_x^s - h_x = 0, \quad \forall x \in X_N;$

(3) $h_x^s = h_y^s, \quad \forall x \in X_C, y = s(x);$

(4) $h_x^s = \sum_{y \in X(x)} p_{x,y} h_y^s, \quad \forall x \in X_N.$

If h_x^s for $x \in X$ satisfies the conditions

$$h_x^s \le h_y^s \text{ for } x \in X_C, y = s(x) \text{ and } h_x^s \le \sum_{y \in X(x)} p_{x,y} h_y^s \text{ for } x \in X_N$$

then

$$c_{x,y} + \varepsilon_y^s - \varepsilon_x^s - h_x^s \ge 0, \quad \forall x \in X_C, \ y = s(x);$$

$$\mu_x + \sum_{y \in X(x)} p_{x,y}\varepsilon_y^s - \varepsilon_x^s - h_x \ge 0, \quad \forall x \in X_N.$$

If h_x^s for $x \in X$ satisfies the conditions

$$h_x^s \ge h_y^s \text{ for } x \in X_C, y = s(x) \text{ and } h_x^s \ge \sum_{y \in X(x)} p_{x,y} h_z^s \text{ for } x \in X_N$$

then

$$c_{x,y} + \varepsilon_y^s - \varepsilon_x^s - h_x^s \le 0, \quad \forall x \in X_C, \ y = s(x);$$
$$\mu_x + \sum_{y \in X(x)} p_{x,y}\varepsilon_y^s - \varepsilon_x^s - h_x \le 0, \quad \forall x \in X_N.$$

If in Lemma 2.18 we vary the strategy s then as a consequence from this lemma we obtain the following result.

Lemma 2.20 *Let (G, X_C, X_N, c, p) be an arbitrary decision network. Then there exist a function $\varepsilon^* : X \to \mathbb{R}$ and the values h_x^* for $x \in X$ such that for an arbitrary stationary strategy s of the control problem on the network it holds*

$$\begin{cases} \mu_x^s + \sum_{y \in X} p_{x,y}^s \varepsilon_y^* - \varepsilon_x^* - h_x^* \ge 0, & \forall x \in X; \\ h_x^* - \sum_{y \in X(x)} p_{x,y}^s h_y^* \le 0, & \forall x \in X; \end{cases} \tag{2.48}$$

Moreover, there exists a stationary strategy s^* such that

$$h_x^* - \sum_{y \in X(x)} p_{x,y}^{s^*} h_y^* = 0, \qquad \forall x \in X, \tag{2.49}$$

where ε_x^* for $x \in X$ represents a solution of the system of the equation

$$\mu_x^{s^*} + \sum_{y \in X} p_{x,y}^{s^*} \varepsilon_y - \varepsilon_x - h_x^* = 0, \quad \forall x \in X. \tag{2.50}$$

Corollary 2.21 *For an arbitrary decision network (G, X_C, X_N, c, p) there exist the values ε_x^*, h_x^* for $x \in X$ that represent the solution of the system of linear inequalities*

$$\begin{cases} c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - h_x^* \geq 0, & \forall x \in X_C, \quad \forall y \in X(x); \\ \mu_x + \sum\limits_{y \in X(x)} p_{x,y} \varepsilon_y^* - \varepsilon_x^* - h_x^* \geq 0, & \forall x \in X_N; \\ h_x^* - h_y^* \leq 0, & \forall x \in X_C, \ \forall y \in X(x); \\ h_x^* - \sum\limits_{y \in X(x)} p_{x,y} h_y^* \leq 0, & \forall x \in X_N, \end{cases} \tag{2.51}$$

where h_x^* for $x \in X$ satisfy the condition

$$\begin{cases} \min\limits_{y \in X(x)} \{h_y^* - h_x^*\} = 0, & \forall x \in X_C; \\ h_x^* - \sum\limits_{y \in X(x)} p_{x,y} h_y^* = 0, & \forall x \in X_N \end{cases} \tag{2.52}$$

and ε_x^* for $x \in X$ represents a solution of the system of linear equations

$$\begin{cases} c_{x,y} + \varepsilon_y - \varepsilon_x - h_x^* = 0, & \forall (x, y) \in E_{h^*}^*; \\ \mu_x + \sum\limits_{y \in X(x)} p_{x,y} \varepsilon_y - \varepsilon_x - h_x^* = 0, & \forall x \in X_N, \end{cases} \tag{2.53}$$

where $E_{h^*}^* = \left\{(x, y) \in E_C \,|\, x \in X_C, \ y \in \operatorname{argmin}_{z \in X(x)} h_z^*\right\}$.

Proof of Theorem 2.12. According to Lemma 2.20 and Corollary 2.21 for the decision network (G, X_C, X_N, c, p) there exist the function $\varepsilon^* : X \to \mathbb{R}$ and the values h_x^* for $x \in X$ that satisfy the conditions (2.51)–(2.53). Thus, we can determine as a basic solution of the system (2.51) and the potential transformation

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - h_x^*, \quad \forall x \in X_C, \ \forall y \in X(x)$$

that corresponds to the network $(G, X_C, X_N, \overline{c}, p)$ in canonical form. We obtain an optimal stationary strategy s^* for the problem on this network if for every $x \in X$ we fix $s^*(x) = y^*$, where y^* satisfies the condition $\overline{c}_{x,y^*}^{s^*} = 0$. Based on Lemma (2.17) we have

$$0 = \overline{\omega}_x^{s^*} = \omega_x^{s^*} - h_x^*, \quad \forall x \in X,$$

and for an arbitrary other strategy s it holds $\omega_x^s - h^* \geq 0$. So, $\omega_x^* = h_x^*$, $\forall x \in X$.

2.2.7 *Linear Programming for Multichain Control Problems and an Algorithm for Determining Optimal Stationary Strategies*

We develop the linear programming approach for a multichain control problem on the bases of the optimality criterion established in Theorem 2.12 and Corollary 2.21. If the vectors $\overline{h}^*$ and $\overline{\varepsilon}^*$ with the corresponding components h_x^* and ε_x^* satisfy the condition (2.52), then we obtain the vector of optimal average costs ω^*. Consequently we have to determine the "maximal" $\overline{h}^*$ that satisfies (2.51). This means that we have to maximize the positive linear combination of components of $\overline{h}^*$ that satisfy (2.51).

Thus, we can determine ε^* and ω^* if we solve the following linear programming problem:
Maximize

$$\overline{\psi}\,'(\varepsilon, \omega) = \sum_{x \in X} \theta_x \omega_x \tag{2.54}$$

subject to

$$\begin{cases} \varepsilon_x - \varepsilon_y + \omega_x \le c_{x,y}, \forall x \in X_C, & \forall y \in X(x); \\ \varepsilon_x - \sum\limits_{y \in X(x)} p_{x,y}\varepsilon_y + \omega_x \le \mu_x, & \forall x \in X_N; \\ \omega_x - \omega_y \le 0, & \forall x \in X_C,\ \forall y \in X(x); \\ \omega_x - \sum\limits_{y \in X(x)} p_{x,y}\omega_y \le 0, & \forall x \in X_N; \end{cases} \tag{2.55}$$

where $\theta_x > 0,\ \forall x \in X$ and $\sum_{x \in X} \theta_x = 1$.

Note that in this model in the case of the unichain control problem the restrictions $\omega_x - \omega_y \le 0$ for $x \in X_C,\ y \in X(x)$ and $\omega_x - \sum_{y \in X(x)} p_{x,y}\omega_y \le 0$ for $x \in X_N$ become redundant in (2.55), because here we can take $\omega_x = \omega_y,\ \forall x, y \in X$. Thus, this model generalizes the linear programming model (2.21), (2.22). Using this model we can propose the following algorithm for determining the solution of the multichain control problem.

Remark 2.22 In (2.55) the inequalities that correspond to the states $x \in X_N$ can be changed by equalities, i.e., the constraints (2.55) in the problem (2.54), (2.55) can be replaced by the constraints

$$\begin{cases} \varepsilon_x - \varepsilon_y + \omega_x \le c_{x,y}, & \forall x \in X_C,\ \forall y \in X(x); \\ \varepsilon_x - \sum\limits_{y \in X(x)} p_{x,y}\varepsilon_y + \omega_x = \mu_x, & \forall x \in X_N; \\ \omega_x - \omega_y \le 0, & \forall x \in X_C,\ \forall y \in X(x); \\ \omega_x - \sum\limits_{y \in X(x)} p_{x,y}\omega_y = 0, & \forall x \in X_N. \end{cases} \tag{2.56}$$

Thus, the optimal solutions of the problems (2.54), (2.55) and (2.54), (2.56) are the same.

Algorithm 2.23 Determining the Optimal Stationary Strategies for the Multichain Control Problem

(1) Formulate the linear programming problem (2.54), (2.55) and determine an optimal solution ε^*, ω^* that satisfies the conditions (2.52), (2.53).
(2) Formulate the potential transformation

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - \omega_x^*, \quad \forall (x, y) \in E.$$

(3) Determine the set

$$E_{\overline{c}}^*(x) = \left\{ (x, y) \in E_C \;\middle|\; y \in \operatorname*{argmin}_{z \in X(x)} \overline{c}_{x,z} \right\}, \quad \forall x \in X_C;$$

$$E_{\omega^*}^*(x) = \left\{ (x, y) \in E_C \;\middle|\; y \in \operatorname*{argmin}_{z \in X(x)} \omega_z^* \right\}, \quad \forall x \in X_C;$$

(4) Fix a strategy $s^* : X_C \to X$ such that $s^*(x) = y$ for every $x \in X$, where $(x, y) \in E_{\overline{c}^*}(x) \cap E_{\omega^*}^*(x)$.

Below we illustrate Algorithm 2.23 based on the following example.

Example Consider the stochastic control problem on network (G, X_1, X_2, c, p) with the structure of the graph $G = (X, E)$ given in Fig. 2.3.

In this graph the vertices are represented by circles and squares. The vertices represented by circles correspond to the controllable states of the dynamical system and the vertices represented by squares correspond to uncontrollable states.

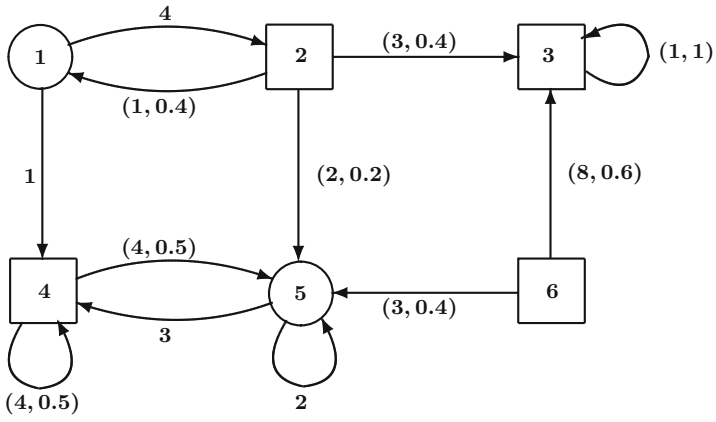

Fig. 2.3 The structure of the graph $G = (X, E)$

So,

$$X = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\};\quad X_C = \{1,\ 5\};\quad X_N = \{2,\ 3,\ 4,\ 6\};$$

$$\begin{aligned} E &= \{(1,2),\ (1,4),\ (2,1),\ (2,3),\ (2,5),\ (3,3),\ (4,4),\ (4,5),\\ &\qquad (5,5),\ (5,4),\ (6,3),\ (6,5)\};\\ E_C &= \{(1,2),\ (1,4),\ (5,5),\ (5,4)\};\\ E_N &= \{(2,1),\ (2,3),\ (2,5),\ (3,3),\ (4,4),\ (4,5),\ (6,3),\ (6,5)\}. \end{aligned}$$

The values of the cost function $c : E \to \mathbb{R}$ and of the transition probability function $p : E \to \mathbb{R}$ are written close to the edges in the picture. For the edges $e = (x, y) \in E_N$ these values are written in parentheses, where the first quantity expresses the cost and the second one represents the probability transition from the state x to the state y. For the edges $e = (x, y) \in E_C$ only the costs are given which are written also close to the edges. Thus, for this example we obtain:

$$\begin{aligned} &c_{1,2} = 4,\quad c_{1,4} = 1,\quad c_{2,1} = 1,\quad c_{2,3} = 3,\quad c_{2,5} = 2,\quad c_{3,3} = 1,\\ &c_{4,4} = 4,\quad c_{4,5} = 4,\quad c_{5,5} = 2,\quad c_{5,4} = 3,\quad c_{6,3} = 8,\quad c_{6,5} = 3;\\ &p_{2,1} = 0.4,\quad p_{2,3} = 0.4,\quad p_{2,5} = 0.2,\quad p_{3,3} = 1,\quad p_{4,4} = 0.5,\\ &p_{4,5} = 0.5,\quad p_{6,3} = 0.6,\quad p_{6,5} = 0.4. \end{aligned}$$

We apply Algorithm 2.23. Afterwards, we solve the linear programming problem: Maximize

$$\overline{\Psi}'(\varepsilon, \omega) = \theta_1\omega_1 + \theta_2\omega_2 + \theta_3\omega_3 + \theta_4\omega_4 + \theta_5\omega_5 + \theta_6\omega_6$$

subject to

$$\begin{cases} \varepsilon_1 - \varepsilon_2 + \omega_1 \le c_{1,2};\\ \varepsilon_1 - \varepsilon_4 + \omega_1 \le c_{1,4};\\ \varepsilon_5 - \varepsilon_4 + \omega_5 \le c_{5,4};\\ \varepsilon_5 - \varepsilon_5 + \omega_5 \le c_{5,5};\\ \varepsilon_2 - (p_{2,1}\varepsilon_1 + p_{2,3}\varepsilon_3 + p_{2,5}\varepsilon_5) + \omega_2 \le \mu_2;\\ \varepsilon_3 - p_{3,3}\varepsilon_3 + \omega_3 \le \mu_3;\\ \varepsilon_4 - (p_{4,4}\varepsilon_4 + p_{4,5}\varepsilon_5) + \omega_4 \le \mu_4;\\ \varepsilon_6 - (p_{6,3}\varepsilon_3 + p_{6,5}\varepsilon_5) + \omega_6 \le \mu_6;\\ \omega_1 - \omega_2 \le 0,\quad \omega_1 - \omega_4 \le 0;\\ \omega_5 - \omega_5 \le 0,\quad \omega_5 - \omega_4 \le 0;\\ \omega_2 - (p_{2,1}\omega_1 + p_{2,3}\omega_3 + p_{2,5}\omega_5) \le 0;\\ \omega_3 - p_{3,3}\omega_3 \le 0;\\ \omega_4 - (p_{4,4}\omega_4 + p_{4,5}\omega_5) \le 0;\\ \omega_6 - (p_{6,3}\omega_3 + p_{6,5}\omega_5) \le 0. \end{cases}$$

Here

$$\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = \theta_6 = \frac{1}{6}$$

and

$$\mu_2 = 2,\ \mu_3 = 1,\ \mu_4 = 4,\ \mu_6 = 6.$$

If we introduce these data in the linear programming model above then we obtain the problem:
Maximize

$$\overline{\Psi}'(\varepsilon, \omega) = \frac{1}{6}\omega_1 + \frac{1}{6}\omega_2 + \frac{1}{6}\omega_3 + \frac{1}{6}\omega_4 + \frac{1}{6}\omega_5 + \frac{1}{6}\omega_6$$

subject to

$$\begin{cases} \varepsilon_1 - \varepsilon_2 + \omega_1 \le 4; \\ \varepsilon_1 - \varepsilon_4 + \omega_1 \le 1; \\ \varepsilon_5 - \varepsilon_4 + \omega_5 \le 3; \\ \omega_5 \le 2; \\ \varepsilon_2 - 0.4\varepsilon_1 - 0.4\varepsilon_3 - 0.2\varepsilon_5 + \omega_2 \le 2; \\ \omega_3 \le 1; \\ \varepsilon_4 - 0.5\varepsilon_4 - 0.5\varepsilon_5 + \omega_4 \le 4; \\ \varepsilon_6 - 0.6\varepsilon_3 - 0.4\varepsilon_5 + \omega_6 \le 6; \\ \omega_1 - \omega_2 \le 0, \quad \omega_1 - \omega_4 \le 0, \quad \omega_5 - \omega_4 \le 0; \\ \omega_2 - 0.4\omega_1 - 0.4\omega_3 - 0.2\omega_5 \le 0; \\ \omega_4 - 0.5\omega_4 - 0.5\omega_5 \le 0; \\ \omega_6 - 0.6\omega_3 - 0.4\omega_5 \le 0. \end{cases}$$

The optimal solution of this problem that satisfies the conditions (2.52), (2.53) is:

$$\varepsilon_1^* = 0,\ \ \varepsilon_2^* = -\frac{8}{3},\ \ \varepsilon_3^* = -\frac{25}{3},\ \ \varepsilon_4^* = 4,\ \ \varepsilon_5^* = 0,\ \ \varepsilon_6^* = -\frac{2}{5};$$
$$\omega_1^* = \frac{4}{3},\ \ \omega_2^* = \frac{4}{3},\ \ \omega_3^* = 1,\ \ \omega_4^* = 2,\ \ \omega_5^* = 2,\ \ \omega_6^* = \frac{7}{5}.$$

If we determine the potential transformation

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - \omega_x^*, \quad \forall (x, y) \in E$$

then we obtain

$$\overline{c}_{1,2} = 0,\ \ \overline{c}_{1,4} = \frac{11}{3},\ \ \overline{c}_{2,1} = \frac{7}{3},\ \ \overline{c}_{2,3} = -4,\ \ \overline{c}_{2,5} = \frac{10}{3},$$
$$\overline{c}_{3,3} = 0,\ \ \overline{c}_{4,5} = -2,\ \ \overline{c}_{4,4} = 2,\ \ \overline{c}_{5,4} = 5,\ \ \overline{c}_{5,5} = 0,\ \ \overline{c}_{6,3} = -\frac{4}{3},\ \ \overline{c}_{6,5} = 2;$$

$$\overline{\mu}_2 = 0,\ \ \overline{\mu}_3 = 0,\ \ \overline{\mu}_4 = 0,\ \ \overline{\mu}_6 = 0.$$

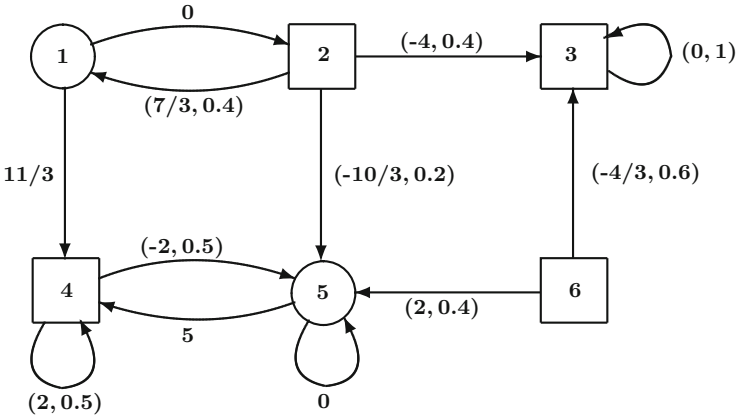

Fig. 2.4 The network $(G, X_C, X_N, \overline{c}, p)$ in canonical form

The network $(G, X_C, X_N, \overline{c}, p)$ in canonical form is represented by Fig. 2.4. This network satisfies the conditions:

(1) $\min\{\overline{c}_{1,1}, \overline{c}_{1,4}\} = 0,\ \min\{\overline{c}_{5,5}, \overline{c}_{5,4}\} = 0;$

(2) $\overline{\mu}_2 = 0,\ \overline{\mu}_3 = 0,\ \overline{\mu}_4 = 0,\ \overline{\mu}_6 = 0.$

For a given optimal solution $\varepsilon_x^*,\ \omega_x^*$ for $x \in X$ we have $E_{h^*}^*(1) = E_{\overline{c}}(1) = \{(1, 2)\}$ and $E_{h^*}^*(5) = E_{\overline{c}}(5) = \{(5, 5)\}$.

Therefore, if we fix $s^*(1) = 2$; $s^*(5) = 5$ then we obtain the optimal stationary strategy $s^* : 1 \to 2;\ 5 \to 5$. The corresponding network induced by the optimal stationary strategy s^* is represented by Fig. 2.5.

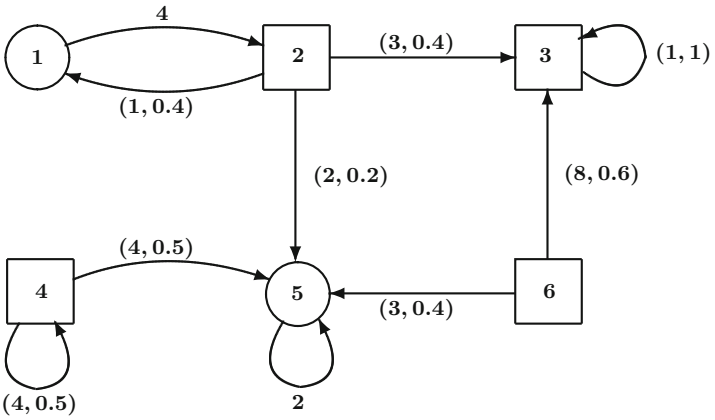

Fig. 2.5 The network induced by the optimal strategy

Another optimal solution of the linear programming problem for this example is:

$$\varepsilon_1^* = 0,\ \varepsilon_2^* = -\frac{8}{3},\ \varepsilon_3^* = -\frac{13}{2},\ \varepsilon_4^* = \frac{1}{3},\ \varepsilon_5^* = -\frac{11}{3},\ \varepsilon_6^* = -\frac{7}{15};$$
$$\omega_1^* = \frac{4}{3},\ \omega_2^* = \frac{4}{3},\ \omega_3^* = 1,\ \omega_4^* = 2,\ \omega_5^* = 2,\ \omega_6^* = \frac{7}{5}.$$

If we calculate $\overline{c}_{x,y}$ and $\overline{\mu}_x$ that correspond to this optimal solution then we obtain

$$\overline{c}_{1,2} = 0,\ \overline{c}_{1,4} = 0,\ \overline{c}_{5,4} = 3,\ \overline{c}_{5,5} = 0,\ \overline{\mu}_2 = 0,\ \overline{\mu}_3 = 0,\ \overline{\mu}_4 = 0,\ \overline{\mu}_6 = 0.$$

It is easy to observe that in this case $E_{\overline{c}}^*(x) \neq E_{h^*}^*(x)$ for $x = 1$. However, we can determine the optimal solution $s^*(1) = 2,\ s^*(5) = 5$ if we fix the strategy s^* such that $(x, s^*(x)) \in E_{\overline{c}}^*(x) \cap E_{h^*}^*(x)$ for $x = 1$ and $x = 2$, i.e., we obtain the same optimal stationary strategy as in the previous case.

Remark 2.24 If for a multichain control problem it is necessary to determine the optimal stationary strategy s^* only for a fixed starting state x_0 then it is sufficient to solve the linear programming problem:
Maximize

$$\overline{\psi}\,'(\varepsilon, \omega) = \omega_{x_0} \tag{2.57}$$

subject to (2.54). The optimal strategy for the considered problem can be found using Algorithm 2.23 if in the item 1 we exchange the problem (2.54), (2.55) by the problem (2.55), (2.57).

If in the example above we fix $x_0 = 1$ and solve the linear programming problem (2.55), (2.57) then we obtain the optimal solution $\varepsilon^*,\ \omega^*$, where

$$\varepsilon_1^* = 0,\ \varepsilon_2^* = -\frac{8}{3},\ \varepsilon_3^* = -\frac{25}{3},\ \varepsilon_4^* = 4,\ \varepsilon_5^* = 0;$$
$$\omega_1^* = \frac{4}{3},\ \omega_2^* = \frac{4}{3},\ \omega_3^* = 1,\ \omega_4^* = 2,\ \omega_5^* = 2$$

and $\varepsilon_6^*,\ \omega_6^*$ are arbitrary values that satisfy the conditions

$$\varepsilon_3^* - 0.6\varepsilon_3^* - 0.4\varepsilon_5^* + \omega_6^* \leq 6,\ \ \omega_6^* - 0.6\omega_3^* - 0.4\omega_5^* \leq 0.$$

Here ε_6^* may differ from $-2/5$ and ω_6^* may differ from $7/5$. In this case we obtain the same optimal strategy $s^* : 1 \to 2;\ 5 \to 5$ but we do not obtain ε_6^* and ω_6^*. If we solve the problem (2.55), (2.57) for $x_0 = 6$ then we obtain

$$\varepsilon_6^* = -\frac{25}{3},\ \omega_6^* = \frac{7}{5},\ \varepsilon_3^* = 0,\ \omega_3^* = 1,\ \varepsilon_5^* = 0,\ \omega_5^* = 2.$$

The remaining variables may be arbitrary.

2.2.8 Primal and Dual Linear Programming Models for the Multichain Problem

The problem (2.54), (2.55) generalizes the unichain dual linear programming model (2.21), (2.22). Therefore, we can regard (2.54), (2.55) as the dual problem of a primal multichain linear programming model. If we dualize (2.54), (2.55) then we obtain a problem which generalizes the problem (2.16), (2.17). This problem can be formulated as follows:
Minimize

$$\overline{\psi}(\alpha, \beta, \lambda, q) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y} \alpha_{x,y} + \sum_{z \in X_N} \mu_z q_z \tag{2.58}$$

subject to

$$\begin{cases} \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - \sum\limits_{x \in X(y)} \alpha_{y,x} + \sum\limits_{x \in X_N} p_{x,y} q_x = 0, \quad \forall y \in X_C; \\ \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - q_y + \sum\limits_{x \in X_N} p_{x,y} q_x = 0, \quad \forall y \in X_N; \\ \sum\limits_{y \in X(x)} \alpha_{x,y} + \sum\limits_{y \in X} \beta_{x,y} - \sum\limits_{y \in X_C^-(x)} \beta_{y,x} - \sum\limits_{y \in X_N^-(x)} p_{y,x} \lambda_y = \theta_x, \forall x \in X_C; \\ q_x + \lambda_x - \sum\limits_{y \in X_N^-(x)} p_{y,x} \lambda_x = \theta_x, \ \forall x \in X_N; \\ \alpha_{x,y}, \beta_{x,y} \geq 0, \ \ \forall x \in X_C, \ y \in X(x); \ \ q_x, \lambda_x \geq 0, \ \ \forall x \in X_N. \end{cases} \tag{2.59}$$

It is easy to see that this linear programming model generalizes the unichain linear programming model (2.16), (2.17). The last two restrictions (equalities) in (2.59) generalize the constraint

$$\sum_{x \in X_C} \sum_{y \in X(x)} \alpha_{x,y} + \sum_{x \in X_N} q_x = 1.$$

In the following we shall regard the linear programming problem (2.54), (2.55).

2.2.9 An Algorithm for Solving the Multichain Control Problem Using a Dual Unichain Model

For multichain control problems the optimal average costs in different states may be different. Therefore, the set of states X can be divided into several subsets $X_1, X_2, \ldots, X_k$ such that each subset X_i, $i \in \{1, 2, \ldots, k\}$ contains the states

with the same optimal average costs and there are no states from different subsets with the same optimal average costs.

Let ω^i be the corresponding optimal average cost of the states $x \in X_i$, $i = 1, 2, \ldots, k$ and assume that $\omega^1 < \omega^2 < \cdots < \omega^k$. In this section we show that the average costs ω^i and the corresponding subsets X_i can be found successively by solving k unichain linear programming problems (2.21), (2.22).

At the first step of the algorithm we solve the linear programming problem: Maximize

$$\overline{\psi}\,'(\varepsilon, h) = h \tag{2.60}$$

subject to

$$\begin{cases} \varepsilon_x - \varepsilon_y + h \le c_{x,y}, & \forall x \in X_C,\ y \in X(x); \\ \varepsilon_x - \sum\limits_{z \in X} p_{x,z}\varepsilon_z + h \le \mu_x, & \forall x \in X_N. \end{cases} \tag{2.61}$$

Let ε_x^1 $(x \in X)$, h^1 be an optimal solution of this problem on the network (G, X_C, X_N, c, p). Then this solution satisfies the conditions:

(1) $c_{x,y}^1 = c_{x,y} + \varepsilon_y^1 - \varepsilon_x^1 - h^1 \ge 0, \qquad \forall x \in X_C,\ y \in X(x);$

(2) $\mu_x^1 = \mu_x + \sum\limits_{y \in X(x)} p_{x,y}\varepsilon_y^1 - \varepsilon_x^1 - h^1 \ge 0, \qquad \forall x \in X_N;$

(3) There exists a nonempty subset X_1 from X where
$\min\limits_{y \in X(x)} c_{x,y}^1 = \min\limits_{y \in X_1(x)} c_{x,y}^1 = 0,\ \forall x \in X_1 \cap X_C;$
$\mu_x^1 = 0,\ \forall x \in X_1 \cap X_N,$
and X_1 is a maximal subset in X with such a property.

If in the network (G, X_C, X_N, c, p) we make the potential transformation

$$c_{x,y}^1 = c_{x,y} + \varepsilon_y^1 - \varepsilon_x^1 - h^1, \qquad \forall x \in X,\ y \in X(x)$$

then we obtain the network (G, X_C, X_N, c^1, p) with a new cost function c^1 on E. According to Lemma 2.14 and Corollary 2.16 the optimal stationary strategies of the control problem on this network are the same as the optimal stationary strategies on the network (G, X_C, X_N, c, p). Moreover, we have here

$$\overline{\omega}_x^1 = \omega_x - h^1, \quad \forall x \in X,$$

where ω_x for $x \in X$ represents the corresponding optimal average costs of the states $x \in X$ in the primal problem and $\overline{\omega}_x^1$ are the optimal average costs of the states in the control problem on the network with transformation potential function c^1.

Thus, after the first step of the algorithm we obtain the subset X_1, the value of the optimal average cost $\omega^1 = h^1$ for the states $x \in X_1$, the function $\varepsilon^1 : X \to \mathbb{R}$

and the network (G, X_C, X_N, c^1, p) with a new cost function c^1, where the optimal average costs $\overline{\omega}_x^1$ in the problem with the new network satisfy the condition:

$$\overline{\omega}_x^1 = 0, \quad \forall x \in X_1; \quad \overline{\omega}_x^1 = \omega_x - h^1 > 0, \quad \forall x \in X \backslash X_1.$$

At the second step of the algorithm we solve the linear programming problem: Minimize the objective function (2.60) subject to

$$\begin{cases} \varepsilon_x - \varepsilon_y + h \le c_{x,y}^1, & \forall x \in X_C \backslash X_1, \ y \in X(x); \\ \varepsilon_x - \sum\limits_{z \in X} p_{x,z}\varepsilon_z + h \le \mu_x^1, & \forall x \in X_N \backslash X_1; \\ \varepsilon_x - \varepsilon_y \le c_{x,y}^1, & \forall x \in X_1 \cap X_C, \ y \in X(x); \\ \varepsilon_x - \sum\limits_{z \in X)} p_{x,z}\varepsilon_z \le \mu_x^1, & \forall x \in X_1 \cap X_N. \end{cases} \tag{2.62}$$

This system is obtained from (2.61) by changing $c_{x,y}$ and μ_x by $c_{x,y}$ and μ_x, and setting $h = 0$ in the inequalities that correspond to the states $x \in X_1$.

Let ε_x^2 $(x \in X)$, h^2 be an optimal solution of this problem on the network (G, X_C, X_N, c^1, p). Then this solution satisfies the conditions:

(1) $c_{x,y}^2 = c_{x,y} + \varepsilon_y^2 - \varepsilon_x^2 - h^2 \ge 0, \qquad \forall x \in X_C, \ y \in X(x);$

(2) $\mu_x^2 = \mu_x + \sum\limits_{y \in X(x)} p_{x,y}\varepsilon_y^2 - \varepsilon_x^2 - h^2 \ge 0, \qquad \forall x \in X_N;$

(3) There exists a nonempty subset X_2 from X where

$$\min_{y \in X(x)} c_{x,y}^2 = \min_{y \in X_2(x)} c_{x,y}^2 = 0, \ \forall x \in X_2 \cap X_C;$$

$$\mu_x^2 = 0, \ \forall x \in X^2 \cap X_N,$$

and X_2 is a maximal subset in X with such a property.

After that we make the potential transformation

$$c_{x,y}^2 = c_{x,y}^1 + \varepsilon_y^2 - \varepsilon_x^2 - h^2, \qquad \forall x \in X, \ y \in X(x)$$

in the network (G, X_C, X_N, c^1, p) and we obtain the network (G, X_C, X_N, c^2, p) with a new cost function c^2 on E. According to Lemma 2.14 and Corollary 2.16 the optimal stationary strategies of the control problem on this network are the same as the optimal stationary strategies on the network (G, X_C, X_N, c^1, p). Moreover, here we have

$$\overline{\omega}_x^2 = \overline{\omega}_x^1 - h^2, \quad \forall x \in X \backslash X_1,$$

where $\overline{\omega}_x^1$ for $x \in X \backslash X_1$ represent the corresponding optimal average costs of the states in the problem before the potential transformation is made and $\overline{\omega}_x^2$ are the

optimal average costs of the states $x \in X \backslash X_1$ in the control problem after the potential transformation is made.

Thus, after the second step of the algorithm we obtain the subset X_2, the value of the optimal average cost h^2 for the states $x \in X_2$, the function $\varepsilon^2 : X \to \mathbb{R}$ and the network (G, X_C, X_N, c^2, p) with a new cost function c^2, where for the optimal average costs $\overline{\omega}_x^2$ in the problem we may set:

$$\overline{\omega}_x^2 = 0, \ \ \forall x \in X_1 \cup X_2; \quad \overline{\omega}_x^2 = \overline{\omega}_x^1 - h^2 > 0, \ \ \forall x \in X \backslash (X_1 \cup X_2).$$

At the next step of the algorithm we solve the linear programming problem: Minimize the objective function (2.60) subject to

$$\begin{cases} \varepsilon_x - \varepsilon_y + h \le c_{x,y}^2, & \forall x \in X_C \backslash (X_1 \cup X_2), \ y \in X(x); \\ \varepsilon_x - \sum\limits_{z \in X} p_{x,z} \varepsilon_z + h \le \mu_x^2, & \forall x \in X_C \backslash (X_1 \cup X_2); \\ \varepsilon_x - \varepsilon_y \le c_{x,y}^2, & \forall x \in (X_1 \cup X_2) \cap X_C; \\ \varepsilon_x - \sum\limits_{z \in X} p_{x,z} \varepsilon_z \le \mu_x^2, & \forall x \in (X_1 \cup X_2) \cap X_N. \end{cases} \tag{2.63}$$

This system is obtained from (2.62) by exchanging $c_{x,y}^1$ and μ_x^1 by $c_{x,y}^2$ and μ_x^2, and setting $h = 0$ in the inequalities that corresponds to the states $x \in X_2$.

After a finite number of steps we obtain the subsets

$$X_1, X_2, \ldots, X_k \ \ (X = X_1 \cup X_2 \cup \cdots \cup X_k),$$

the potential functions $\varepsilon^i : X \to \mathbb{R}, \quad i = 1, 2, \ldots, k$ and the values $h^1, h^2, \ldots, h^k$, where

$$\omega^i = \sum_{j=1}^{i} h^j, \quad j = 1, 2, \ldots, k.$$

If we find $\varepsilon_x^* = \sum_{i=1}^{k} \varepsilon_x^i$ and fix $\omega_x^* = \omega^{i^*}$ for $x \in X_i$ then we determine the potential transformation

$$\overline{c}_{x,y} = c_{x,y} + \varepsilon_y^* - \varepsilon_x^* - \omega_x^*, \ \ \forall x \in X, \ y \in X(x),$$

that satisfies the conditions (1) – (6) of Theorem 2.12. This means that we determine the network $(G, X_C, X_N, \overline{c}, p)$ and the optimal stationary strategy s^*.

Example Consider the stochastic control problem on the network with the data from the example given in the previous section. The network is represented by Fig. 2.3, where $X = X_C \cup X_N, \ X_C = \{1, 5\}, \ X_N = \{2, 3, 4, 6\}$, and the costs and transition probabilities are written again along the edges.

We apply the algorithm described above. At the first step of the algorithm we solve the linear programming problem:
Minimize

$$\overline{\psi}'(\varepsilon, h) = h$$

subject to

$$\begin{cases} \varepsilon_1 - \varepsilon_2 + h \le 4; \\ \varepsilon_1 - \varepsilon_4 + h \le 1; \\ \varepsilon_5 - \varepsilon_4 + h \le 3; \\ \varepsilon_5 - \varepsilon_5 + h \le 2; \\ \varepsilon_2 - 0.4\varepsilon_1 - 0.4\varepsilon_3 - 0.2\varepsilon_5 + h \le 2; \\ \varepsilon_3 - \varepsilon_3 + h \le 1; \\ \varepsilon_4 - 0.5\varepsilon_4 - 0.5\varepsilon_5 + h \le 4; \\ \varepsilon_6 - 0.6\varepsilon_3 - 0.4\varepsilon_5 + h \le 6. \end{cases}$$

An optimal solution of this problem is $h^1 = 1$, $\varepsilon_1^1 = 0$, $\varepsilon_2^1 = 0$, $\varepsilon_3^1 = 0$, $\varepsilon_4^1 = 0$, $\varepsilon_5^1 = 0$, $\varepsilon_6^1 = 0 = 1$. We calculate $c_{x,y}^1$ and μ_x^1 using the formula

$$c_{x,y}^1 = c_{x,y} + \varepsilon_y^1 - \varepsilon_x^1 - h^1, \ \forall x \in X_1, \ y \in X(x);$$
$$\mu_x^1 = \mu_x + \sum_{y \in X(x)} p_{x,y} c_{x,y} - \varepsilon_x^1 - h^1, \ x \in X_2$$

and determine $c_{1,2}^1 = 3$, $c_{1,4}^1 = 0$, $c_{5,4}^1 = 2$, $c_{5,5}^1 = 1$; $\mu_2^1 = 1$, $\mu_3^1 = 0$, $\mu_4^1 = 3$, $\mu_6^1 = 5$. After the first step of the algorithm we obtain:

$$X_1 = \{3\}; \ h^1 = 1; \ \varepsilon_1^1 = 0, \ \varepsilon_2^1 = 0, \ \varepsilon_3^1 = 0, \ \varepsilon_4^1 = 0, \ \varepsilon_5^1 = 0, \ \varepsilon_6^1 = 0.$$

At the second step of the algorithm we solve the linear programming problem:
Minimize

$$\overline{\psi}'(\varepsilon, h) = h$$

subject to

$$\begin{cases} \varepsilon_1 - \varepsilon_2 + h \le 3; \\ \varepsilon_1 - \varepsilon_4 + h \le 0; \\ \varepsilon_5 - \varepsilon_4 + h \le 2; \\ \varepsilon_5 - \varepsilon_5 + h \le 1; \\ \varepsilon_3 - \varepsilon_3 \le 0; \\ \varepsilon_2 - 0.4\varepsilon_1 - 0.4\varepsilon_3 - 0.2\varepsilon_5 + h \le 1; \\ \varepsilon_4 - 0.5\varepsilon_4 - 0.5\varepsilon_5 + h \le 3; \\ \varepsilon_6 - 0.6\varepsilon_3 - 0.4\varepsilon_5 + \omega \le 5. \end{cases}$$

An optimal solution of this problem is

$$h^2 = \frac{1}{3},\ \varepsilon_1^2 = 0,\ \varepsilon_2^2 = -\frac{8}{3},\ \varepsilon_3^2 = -\frac{25}{3},\ \varepsilon_4^2 = 4,\ \varepsilon_5^2 = 0,\ \varepsilon_6^2 = -\frac{2}{5}.$$

We calculate $c_{x,y}^2$ and μ_x^2 using formula

$$c_{x,y}^2 = c_{x,y}^1 + \varepsilon_y^2 - \varepsilon_x^2 - h^2;\ \ \mu_x^2 = \mu_x^1 + \sum_{z \in X} p_{x,y}\varepsilon_z - h^2$$

and find

$$c_{1,2}^2 = 0,\ c_{1,4}^2 = \frac{2}{3},\ c_{5,4}^2 = \frac{17}{3},\ c_{5,5}^2 = \frac{2}{3},\ \mu_2^2 = 0,\ \mu_3^2 = 0,\ \mu_4^2 = \frac{2}{3},\ \mu_6^2 = \frac{1}{15}.$$

After the second step of the algorithm we obtain: $X_2 = \{1, 2\}$;

$$h^2 = \frac{1}{3},\ \ \varepsilon_1^2 = 0,\ \ \varepsilon_2^2 = -\frac{8}{3},\ \ \varepsilon_3^2 = -\frac{25}{3},\ \ \varepsilon_4^2 = 4,\ \ \varepsilon_5^2 = 0,\ \ \varepsilon_6^2 = -\frac{2}{5}.$$

At the third step of the algorithm we solve the linear programming problem: Minimize

$$\overline{\psi}'(\varepsilon, h) = h$$

subject to

$$\begin{cases} \varepsilon_1 - \varepsilon_2 \le 0; \\ \varepsilon_1 - \varepsilon_4 \le \dfrac{11}{3}; \\ \varepsilon_5 - \varepsilon_4 + h \le \dfrac{17}{3}; \\ \varepsilon_3 - \varepsilon_3 \le 0; \\ \varepsilon_5 - \varepsilon_5 + h \le \dfrac{2}{3}; \\ \varepsilon_2 - 0.4\varepsilon_1 - 0.4\varepsilon_3 - 0.2\varepsilon_5 \le \dfrac{2}{3}; \\ \varepsilon_4 - 0.5\varepsilon_4 - 0.5\varepsilon_5 + h \le \dfrac{2}{3}; \\ \varepsilon_6 - 0.6\varepsilon_3 - 0.4\varepsilon_5 + h \le \dfrac{1}{15}. \end{cases}$$

An optimal solution of this problem is

$$h^3 = \frac{1}{15},\ \varepsilon_1^3 = 0,\ \varepsilon_2^3 = 0,\ \varepsilon_3^3 = 0,\ \varepsilon_4^3 = 0,\ \varepsilon_5^3 = 0,\ \varepsilon_6 = 0.$$

Using this solution we find

$$c^3_{1,2} = 0,\ c^4_{1,4} = \frac{11}{3},\ c^3_{5,5} = \frac{3}{5},\ c^3_{5,4} = \frac{26}{5},\ \mu^3_2 = 0,\ \ \mu^3_3 = 0,\ \mu^3_4 = \frac{3}{5},\ \mu^3_6 = 0.$$

After this step we obtain:

$$X_3 = \{6\};\ h^3 = \frac{1}{15},\ \varepsilon^3_1 = 0,\ \varepsilon^3_2 = 0,\ \varepsilon^3_3 = 0,\ \varepsilon^3_4 = 0,\ \varepsilon^3_5 = 0,\ \varepsilon_6 = 0.$$

At the fourth step of the algorithm we solve the linear programming problem: Minimize

$$\overline{\psi}\,'(\varepsilon, h) = h$$

subject to

$$\begin{cases} \varepsilon_1 - \varepsilon_2 \le 0; \\ \varepsilon_1 - \varepsilon_4 \le \dfrac{11}{3}; \\ \varepsilon_5 - \varepsilon_4 + h \le \dfrac{28}{5}; \\ \varepsilon_3 - \varepsilon_3 \le 0; \\ \varepsilon_5 - \varepsilon_5 + h \le \dfrac{3}{5}; \\ \varepsilon_2 - 0.4\varepsilon_1 - 0.4\varepsilon_3 - 0.2\varepsilon_5 \le \dfrac{2}{3}; \\ \varepsilon_4 - 0.5\varepsilon_4 - 0.5\varepsilon_5 + h \le \dfrac{3}{5}; \\ \varepsilon_6 - 0.6\varepsilon_3 - 0.4\varepsilon_5 \le 0. \end{cases}$$

An optimal solution of this system is $h^4 = 3/5$, $\varepsilon^3_1 = 0$, $\varepsilon^3_2 = 0$, $\varepsilon^3_3 = 0$, $\varepsilon^3_4 = 0$, $\varepsilon^3_5 = 0$, $\varepsilon_6 = 0$. Using this solution we find $c^4_{1,2} = 0$, $c^4_{2,4} = 11/3$, $c^4_{5,4} = 5$, $c^4_{5,5} = 0$, $\mu^4_2 = 0$, $\mu^4_3 = 0$, $\mu^4_4 = 0$, $\mu^4_6 = 0$. After this step we obtain $X_4 = \{4, 5\}$ and $h^4 = 3/5$.

Thus, finally we have $X = X_1 \cup X_2 \cup X_3 \cup X_4$, where

$$X_1 = \{3\},\ X_2 = \{1, 2\},\ X_3 = \{6\},\ X_4 = \{4, 5\},$$

and

$$\omega^1 = h^1,\ \ \omega^2 = h^1 + h^2,\ \ \omega 3 = h^1 + h^2 + h^3,\ \ \omega^4 = h^1 + h^2 + h^3 + h^4,$$

i.e.,

$$\omega^1 = 1,\ \omega^2 = \frac{4}{3},\ \omega^3 = \frac{7}{5},\ \omega^4 = 2.$$

In addition we can find

$$\varepsilon_1^* = \varepsilon_1^1 + \varepsilon_1^2 + \varepsilon_1^3 + \varepsilon_1^4 = 0; \quad \varepsilon_2^* = \varepsilon_2^1 + \varepsilon_2^2 + \varepsilon_2^3 + \varepsilon_2^4 = -\frac{8}{3};$$
$$\varepsilon_3^* = \varepsilon_3^1 + \varepsilon_3^2 + \varepsilon_3^3 + \varepsilon_3^4 = -\frac{25}{3}; \quad \varepsilon_4^* = \varepsilon_4^1 + \varepsilon_4^2 + \varepsilon_4^3 + \varepsilon_4^4 = 4;$$
$$\varepsilon_5^* = \varepsilon_5^1 + \varepsilon_5^2 + \varepsilon_5^3 + \varepsilon_5^4 = 0; \quad \varepsilon_6^* = \varepsilon_6^1 + \varepsilon_6^2 + \varepsilon_6^3 + \varepsilon_6^4 = -\frac{2}{5}.$$

If we make the potential transformation of the cost function c for ω^* and ε^* found above then we obtain the network in canonical form $(G, X_C, X_N, \overline{c}, p)$ represented by Fig. 2.4 that gives the optimal stationary strategies.

2.2.10 An Approach for Solving the Multichain Control Problem Using a Reduction Procedure to a Unichain Problem

We consider the stochastic control problem on the network (G, X_C, X_N, c, p, x_0) with fixed starting state x_0 and describe an approximation algorithm for determining the optimal solutions which is based on a reduction procedure of the multichain problem to the unichain case.

We describe the reduction procedure in the case if the graph G satisfies the condition that for an arbitrary vertex $x \in X_C$ each outgoing directed edge $e = (x, y)$ ends in X_N, i.e., we assume that

$$E_C = \{e = (x, y) \in E \mid x \in X_C,\ y \in X_N\}.$$

If the graph G does not satisfy this condition then the considered control problem can be reduced to a similar control problem on an auxiliary network $(G', X'_C, X'_N, c', p', x_0)$, where the graph G' satisfies the condition mentioned above. Graph $G' = (X', E')$ is obtained from $G = (X, E)$, where each directed edge $e = (x, y) \in E_C$ is changed by the following two directed edges $e^1 = (x, x_e)$ and $e^2 = (x_e, y)$.

We include each vertex x_e in X'_N and to each edge $e' = (x_e, y)$ we associate the cost $c'_{x_e,y} = c_{x,y}$ and the transition probability $p'_{x_e,y} = 1$. To the edges $e' = (x, x_e)$ we associate the cost $c'_{x,x_e} = c_{(x,y)}$, where $e = (x, y)$. For the edges $e \in E_N$ in the new network we preserve the same costs and transition probabilities as in the initial network, i.e., the cost function c' on E_N and on the set of edges (x, x_e) for $x \in X_C,\ e \in E_C$ is induced by the cost function c. Thus, in the auxiliary network the graph G' is determined by the set of vertices $X' = X'_C \cup X'_N$ and the set of edges $E' = E'_C \cup E'_N$, where $X'_C = X_C;\ X'_N = X_N \cup \{x_e,\ e \in E_C\};$ $E'_C = \{e' = (x, x_e) \mid x \in X_C, e = (x, y) \in E_C\};\ E'_N = E_N \cup \{e' = (x_e, y) \mid e = (x, y) \in E_C, y \in X\}$. It is evident that there exists a bijective mapping between the set of strategies in the states $x \in X_C$ of the network (G, X_C, X_N, c, p, x_0) and the set of strategies in the states $x \in X_C$ of the network $(G', X'_C, X'_N, c', p', x_0)$ that preserves the average costs of the problems on the corresponding networks.

Thus, without loss of generality we may consider that G possesses the property that for an arbitrary vertex $x \in X_C$ each outgoing directed edge $e = (x, y)$ ends in X_N. Additionally, let us assume that the vertex x_0 in G is reachable from every vertex $x \in X_N$. Then an arbitrary strategy s in the considered problem induces a transition probability matrix $P^s = (p^s_{x,y})$ that corresponds to a Markov unichain with a positive recurrent class X^+ that contains the vertex x_0.

Therefore, if we solve the control problem on the network then we obtain the solution of the problem with fixed starting state x_0. So, we obtain such a solution if the network satisfies the condition that for an arbitrary strategy s the vertex x_0 in G_s is attainable for every $x \in X_N$. Now let us assume that this property does not take place. In this case we can reduce our problem to a similar problem on a new auxiliary network $(G'', X''_C, X''_N, p'', c'', x_0)$ for which the property mentioned above holds. This network is obtained from the initial one by the following way: We construct the graph $G'' = (X, E'')$ which is obtained from $G = (X, E)$ by adding new directed edges $e''_{x_0} = (x, x_0)$ from $x \in X_N \backslash \{x_0\}$ to x_0, if for some vertices $x \in X_N \backslash \{x_0\}$ in G there are no directed edges $e = (x, x_0)$ from x to x_0. We define the costs of directed edges $(x, y) \in E''$ in G'' as follows: If $e'' = (x, y) \in E$ then the cost $c''_{e''}$ of this edge in G'' is the same as in G, i.e., $c''_{e''} = c_{e''}$ for $e'' \in E$; if $e'' = (x, x_0) \in E'' \backslash E$ then we put $c''_{e''} = 0$. The probabilities $p''_{x,y}$ for $(x, y) \in E''$ where $x \in X_N$ we define by using the following rule: We fix a small positive value ϵ and put $p''_{x,y} = p_{x,y} - \epsilon p_{x,y}$ if $(x, y) \in E'' \backslash E, \quad y \neq x_0$ and in G there is no directed edge $e = (x, x_0)$ from x to x_0; if in G for a vertex $x \in X \backslash \{x_0\}$ there exists a leaving directed edge $e = (x, x_0)$ then for an arbitrary outgoing directed edge $e = (x, y), \ y \in X(x)$ we put $p''_{x,y} = p_{x,y}$; for the directed edges $(x, x_0) \in E' \backslash E$ we put $p''_{x,x_0} = \epsilon$.

Let us assume that the probabilities $p_{x,y}$ for $(x, y) \in E$ are given in the form of irreducible decimal fractions $p_{x,y} = a_{x,y}/b_{x,y}$.

Additionally, assume that $\epsilon \leq 2^{-2L-2}$, where

$$L = \sum_{(x,y)\in E} \log(a_{x,y} + 1) + \sum_{(x,y)\in E} \log(b_{x,y} + 1) + \sum_{e\in E} \log(|c_e| + 1) + 2\log(n) + 1.$$

Here L is the length of the binary-coded data of the matrix P and of the cost vector c with integer components; each probability $p_{x,y}$ is given by the integer couple $a_{x,y}$, $b_{x,y}$. Then, based on the results from [57, 58] for our auxiliary optimization problem (with approximated data) we can conclude that the solution of this problem will correspond to the solution of our initial problem.

If we consider the control problem on the auxiliary network $(G', X'_C, X'_N, c', p', x_0)$ then we can observe that an arbitrary optimal basic solution of the linear programming problem (2.13), (2.14) satisfies the condition $q^*_{x_0} > 0$ and therefore we can determine the optimal stationary strategy s'^* for the auxiliary problem using Algorithm 2.5. In addition we can observe that if for our stochastic control problem

on the network there exists the optimal stationary strategy s^* then it coincides with an optimal stationary strategy s'^* of the stochastic control problem on the auxiliary network, i.e., $s^* = s'^*$. Moreover, the optimal values of the objective functions $f_{x_0}(s^*)$ can be obtained from the optimal value of the objective function $f'_{x_0}(s'^*)$ in the auxiliary problem using the approximation procedure. So, to find the optimal solution of the problem on the network (G, X_C, X_N, c, p, x_0) it is necessary to construct the auxiliary network $(G', X'_C, X'_N, c', p', x_0)$ where for each vertex $x \in X'_N$ an arbitrary directed edge $e' = (x, y)$ ends in X_N. Then we construct the network $(G'', X''_C, X''_N, c'', p'', x_0)$ and the auxiliary stochastic optimal control problem on this network. If the optimal stationary strategy s'^* in the auxiliary problem is found then we fix $s^* = s'^*$ on X_C.

Example Consider the multichain control problem on the network (G, X_C, X_N, c, p, x_0) represented by Fig. 2.6. In this network the vertices represented by circles correspond to the controllable states of the dynamical system and the vertices represented by squares correspond to uncontrollable states. To each edge that originates in the vertices which correspond to the controllable states the associated cost is written along the edge. To each edge that originates in the vertices that correspond to uncontrollable states the associated cost and the transition probability are written in parentheses. The starting state x_0 is fixed and it corresponds to vertex 2, i.e., $x_0 = 2$.

For this network we have $X_C = \{1, 4, 6\}$, $X_N = \{2, 3, 5\}$ and there exist two edges $(1, 1)$, $(1, 4)$ which start in X_C and end in X_C. The corresponding network $(G', X'_C, X'_N, c', p', x_0)$ is represented on Fig. 2.7. This network is obtained from the network on Fig. 2.7 by adding two new vertices $1'$ and $1''$ on the edges $(1, 1)$ and $(1, 4)$.

The network $(G'', X''_C, X''_N, c'', p'', x_0)$ is represented in Fig. 2.8. This network is obtained from the network in Fig. 2.7 by adding the directed edges (x, x_0) that start in the vertices $x \in X''_N = \{1', 1'', 5, 3\}$ and end in $x_0 = 2$, where $c_{x,x_0} = 0$

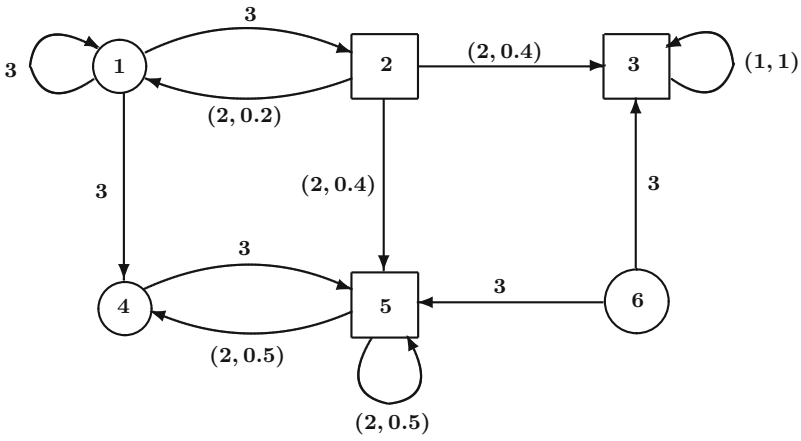

Fig. 2.6 The network (G, X_C, X_N, c, p, x_0)

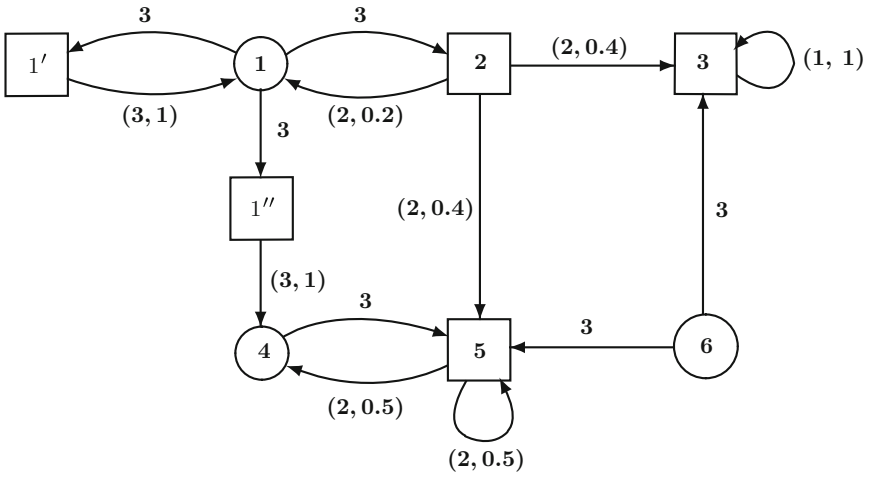

Fig. 2.7 The network $(G', X'_C, X'_N, c', p', x_0)$

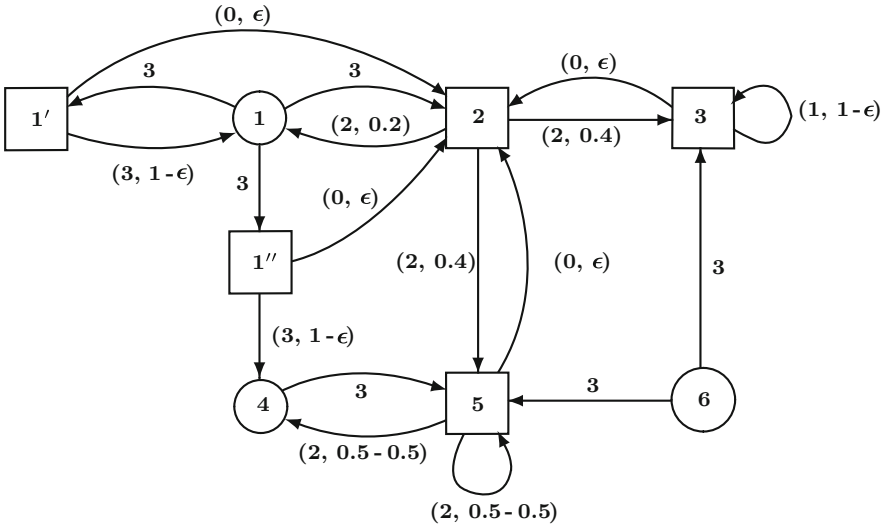

Fig. 2.8 The network $(G'', X''_C, X''_N, c'', p'', x_0)$

and $p_{x,x_0} = \epsilon$. The corresponding probabilities $p_{x,y}$ for the directed edges (x, y) for $x \in X''_N$ are defined as follows: $p''_{x,y} = p_{x,y} - p_{x,y}\epsilon$. The control problem on the auxiliary network possesses the property that an arbitrary strategy s'' generates a Markov unichain. Therefore, for this problem we can use the linear programming model (2.13), (2.14) or the linear programming model (2.21), (2.22) with $\epsilon = 10^{-4}$.

In both cases we determine the same optimal stationary strategy

$$s^{*\prime\prime} : 1 \to 2; \quad 4 \to 5; \quad 6 \to 3.$$

This means that the optimal solution for the initial problem is

$$s^{*} : 1 \to 2; \quad 4 \to 5; \quad 6 \to 3.$$

In the following we show that the linear programming models for the stochastic control problem can be extended for Markov decision processes which lead to the linear programming models from [25, 45, 46, 51, 115].

2.3 A Linear Programming Approach for Markov Decision Problems with an Average Cost Optimization Criterion

We extend now the linear programming approach and algorithms from the previous section for the Markov decision problem with an average cost optimization criterion. We show that an arbitrary Markov decision problem can be transformed into a stochastic control problem on a network and vice versa, an arbitrary stochastic control problem on a 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.

2.3.1 Problem Formulation

A *Markov decision process* [4, 115] 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 satisfies the condition $\sum_{y \in X} p^{a}_{x,y} = 1, \quad \forall a \in A$ and $c : A \times X \times X \to \mathbb{R}$ is a real function. 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 $t = 0, 1, 2, \ldots$. For a fixed action $a \in A$ and arbitrary $x, y \in X$ the function c determines the cost $c^{a}_{x,y}$ of the system's transition from the state $x = x(t)$ to the state $y = x(t + 1)$ for $t = 0, 1, 2, \ldots$. In the considered Markov process the functions p and c do not depend on time, i.e., we have a stationary Markov decision process. If in each state $x \in X$ we fix an action from $a \in A$ then we obtain a Markov process induced by these actions. The problem with an average cost optimization criterion for the Markov decision process (X, A, p, c) with given starting state x_0 consists in determining the actions in the states of the system that provide the minimal (or maximal) average cost per transition for the Markov process

induced by the chosen actions. In the following we will study this problem in terms of stationary strategies.

We define a stationary strategy s for Markov decision process as a map

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

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 the transition cost matrix $C^s = (c^s_{x,y})$. For this Markov process with probability and cost matrices P^s, C^s we can determine the expected average cost per transition $\omega^s_{x_0}$ if the dynamical system starts transitions in the state x_0 at the moment of time $t = 0$. We denote this quantity by $f_{x_0}(s)$, i.e.,

$$f_{x_0}(s) = \omega^s_{x_0}.$$

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

$$f_{x_0}(s^*) = \min_s f_{x_0}(s).$$

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 $x, y \in X$ it 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 $P^{a_k} = (p^{a_k}_{x,y})$, $C^{a_k} = (c^{a_k}_{x,y})$, $\quad k = 1, 2, \ldots, |A|$, where $\sum_{y \in X} p^{a_k}_{x,y} = 1$, $\forall a_k \in A, \forall x \in X$.

A fixed strategy $s : x \to a_k \in A(x)$ for $x \in X$ generates a Markov process with the probability transition matrix P^s and the transition cost matrix C^s induced by the rows of the corresponding matrices P^{a_k} and C^{a_k}, $\quad k = 1, 2, \ldots, |A|$, 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 if $X = X_C \cup X_N$, $X_C \cap X_N = \emptyset$, where for an arbitrary state $x_i \in X_C$ the probabilities $p^{a_k}_{x_i,y}$ are equal to 0 or 1 and for an arbitrary state $x_i \in X_N$ 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 problem can be transformed into a 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. We can observe that the mentioned reduction procedure can be realized in polynomial time. Thus, the considered problems are equivalent from a computational point of view. Using the reduction procedure of the Markov decision problem to a

stochastic control problem we can extend the algorithms from the previous section for determining the optimal solution for Markov decision problems.

2.3.2 Reduction of Markov Decision Problems to Stochastic Control Problems

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 criterion can be reduced to the problem of determining the optimal stationary strategy in the control problem on a network $(G', X'_C, X'_N, p', c', x'_0)$, where $G' = (X', E')$, X'_C, X'_N, p', c' and x_0 are defined in the following way: The set of vertices $X' = X'_C \cup X'_N$ contains $(|A| + 1)|X|$ vertices, where $|X'_C| = |X|$ and $|X'_N| = |A||X|$. So, the set of controllable states in the control problem consists of a copy of the set of states X and the set of uncontrollable states X'_N consists of $|A|$ copies of the set of states X. Therefore, we define X'_C and X_N as follows:

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

where

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

We also represent the set of directed edges E' as a couple of two disjoint subsets $E' = E'_C \cup E'_N$, where E'_C is the set of outgoing edges from $x' \in X'_C$ and E'_N is the set of outgoing edges from $x^a \in X'_N$. The states E'_C and E'_N are defined as follows:

$$E'_C = \{(x, (x, a)) \mid x \in X'_C; \ (x, a) \in X'_N, \ a \in A\};$$
$$E'_N = \{((x, a), y) \mid (x, a) \in X'_N, \ y \in X'_C, \ p^a_{x,y} > 0, \ a \in A\}.$$

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

$$c'_{e'} = 0, \ \ \forall e' = (x, (x, a)) \in E'_C;$$

$$c'_{e'} = 2\, c^a_{x,y} \quad \text{for} \quad e' = ((x, a), y) \in E'_N \quad (x, y \in X, \ a \in A).$$

On E'_2 we define the transition probability function $p' : E'_N \to [0, 1]$, where $p'_{e'} = p^a_{x,y}$ for $e' = ((x, a), y) \in E'_N$.

It is easy to observe that between the set of stationary strategies $\mathbb{S}$ in the Markov decision process and the set of strategies $\mathbb{S}'$ in the control problem on the network $(G', X'_C, X'_N, c', p', x_0)$ there exists a bijective mapping that preserves the average cost per transition. Therefore, if we find the optimal stationary strategy for the control

problem on the network then we can determine the optimal stationary strategy in the 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 using the graph of probability transitions $G_p = (X, E_p)$, which is induced by the probability function $p : X \times X \times A \to [0, 1]$. This graph may contain parallel directed edges where each directed edge corresponds to an action. The set of vertices X corresponds to the set of states and the set of edges E_p consists of $|A|$ subsets $E_p^1, E_p^2, \ldots, E_p^{|A|}$ $\left(E_p = \bigcup_{i=1}^{|A|} E_p^i\right)$, where $E_p^i = \{e^{a_i} = (x, y)^{a_i} \mid p_{x,y}^{a_i} > 0\}$, $i = 1, 2, \ldots, |A(x)|$.

An example how to construct the graph $G_p = (X, E_p)$ and how to determine the solution of the Markov decision problem using the reduction procedure to an auxiliary control problem on the network is given below.

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:

$$
\begin{aligned}
&p_{1,1}^{a_1} = 0.7,\ p_{1,2}^{a_1} = 0.3,\ p_{2,1}^{a_1} = 0.6,\ p_{2,2}^{a_1} = 0.4,\\
&p_{1,1}^{a_2} = 0.4,\ p_{1,2}^{a_2} = 0.6,\ p_{2,1}^{a_2} = 0.5,\ p_{2,2}^{a_2} = 0.5;\\
&c_{1,1}^{a_1} = 1,\ c_{1,2}^{a_1} = 0,\ c_{2,1}^{a_1} = -2,\ c_{2,2}^{a_1} = 5,\\
&c_{1,1}^{a_2} = 0,\ c_{1,2}^{a_2} = 4,\ c_{2,1}^{a_2} = 2,\ c_{2,2}^{a_2} = -3.
\end{aligned}
$$

We consider the problem of finding the optimal stationary strategy for the corresponding Markov decision problem with minimal average costs 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

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

In Fig. 2.9 this Markov process is represented by the multigraph $G_p = (X, E_p)$ with the set of vertices $X = \{1, 2\}$.

The set of directed edges E_p contains parallel directed edges that correspond to probability transitions from one state to another for different actions. We call this graph *multigraph of the Markov decision process*.

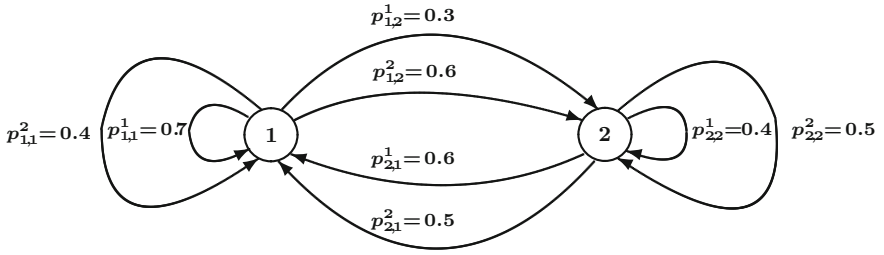

Fig. 2.9 The graph of the Markov decision process

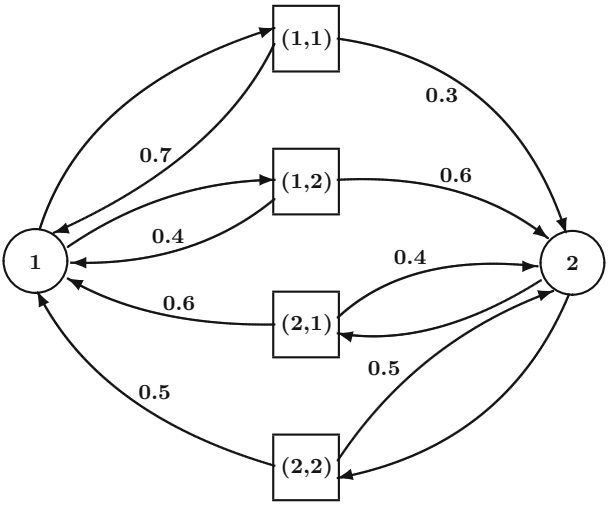

Fig. 2.10 The graph G' for the control problem

In Fig. 2.10 the graph $G' = (X', E')$ is represented. In G' the sets X'_C, X'_N, E'_C, E'_N are defined as follows:

$$X'_C = \{1, 2\},\ X'_N = X^1 \cup X^2 = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$$

where

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

and

$$\begin{aligned} E'_C &= \{(1, (1, 1)),\ (1, (1, 2)),\ (2, (2, 1)),\ (2, (2, 2))\}, \\ E'_N &= \{((1, 1), 1),\ ((1, 1), 2),\ ((2, 1), 1),\ ((2, 2), 1), \\ &\qquad ((1, 2), 1),\ ((1, 2), 2),\ ((2, 1), 2),\ ((2, 2), 2)\}. \end{aligned}$$

The probabilities $p'_e = p'_{(x,a),y} = p^a_{x,y}$ for directed edges $((x,a),y) \in E'_N$ are written along the edges in Fig. 2.10 and the costs of the directed edges from E' are defined in the following way:

$$c'_{1,(1,1)} = c'_{1,(1,2)} = 0, \qquad c'_{2,(2,1)} = c'_{2,(2,2)} = 0,$$

$$\begin{array}{llll} c'_{(1,1),1} = 2, & c'_{(1,1),2} = 0, & c'_{(2,1),1} = -4, & c'_{(2,2),1} = \ \ 4, \\ c'_{(1,2),1} = 0, & c'_{(1,2),2} = 8, & c'_{(2,1),2} = \ 10, & c'_{(2,2),2} = -6. \end{array}$$

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

$$\begin{array}{ll} s^1 : 1 \to a_1, & 2 \to a_1; \\ s^2 : 1 \to a_1, & 2 \to a_2; \\ s^3 : 1 \to a_2, & 2 \to a_1; \\ s^4 : 1 \to a_2, & 2 \to a_2. \end{array}$$

A fixed strategy s in the 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 is easy to check that this Markov process is ergodic and the limit matrix of this process is

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

We can determine the components of the vector of immediate costs $\mu^{s_2} = \begin{pmatrix} \mu_1^{s_2} \\ \mu_2^{s_2} \end{pmatrix}$ 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 $f_1(s_2) = f_2(s_2) = 1/4$. Analogously, it can be calculated by $f_1(s_1) = f_2(s_1) = 22/30$, $f_1(s_3) = f_2(s_3) = 16/10$ and $f_1(s_4) = f_2(4) = 9/11$. We can see that the optimal stationary strategy for the Markov decision problem with minimal average cost criterion is s^2. This strategy can be found by solving the following linear programming problem on the auxiliary network (G', X'_C, X'_N, p', c'):

Minimize

$$\overline{\psi}(\alpha, q) = 1.4q_{1,1} + 4.8q_{1,2} + 1.6q_{2,1} - q_{2,2}$$

subject to

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

The optimal solution of this problem is

$$q_1^* = \frac{5}{16}, \quad q_2^* = \frac{3}{16}, \quad q_{1,1}^* = \frac{5}{16}, \quad q_{2,2}^* = \frac{3}{16}, \quad q_{1,2}^* = 0, \quad q_{2,1}^* = 0,$$
$$\alpha_{1,(1,1)}^* = \frac{5}{16}, \quad \alpha_{2,(2,2)}^* = \frac{3}{16}, \quad \alpha_{1,(1,2)}^* = 0, \quad \alpha_{2,(2,1)}^* = 0$$

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

The optimal strategy s^* on G' we can find using Theorem 2.3, i.e., we fix

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

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

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

and the average cost per transaction 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 in Fig. 2.11. The unique outgoing directed edge $(1, (1, 1))$ from vertex 1 that ends 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 ends in vertex $(2, 2)$ corresponds to the optimal strategy $2 \to a_2$ in the state $x = 2$.

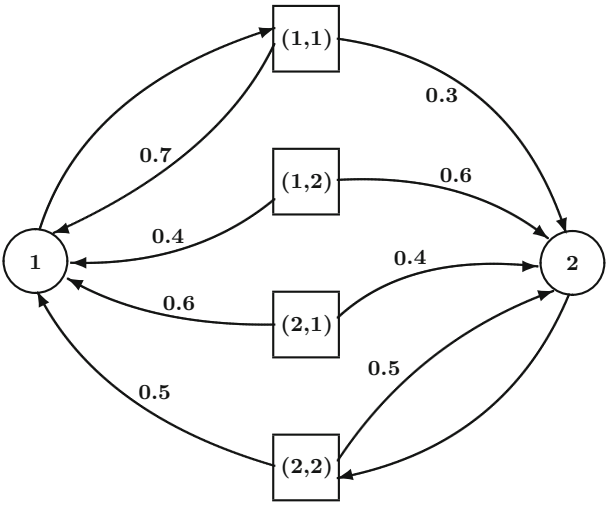

Fig. 2.11 The graph induced by the optimal strategy in the control problem

2.3.3 A Linear Programming Approach for the Average Markov Decision Problem and an Algorithm for Determining the Optimal Strategies

In the previous sections 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 a linear programming algorithm directly to the Markov decision problem with an average cost optimization criterion without constructing 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 $s : X \to A$ generates an ergodic Markov chain, i.e., we assume that the graph $G_p^s = (X, E_p^s)$ of the matrix of probability transitions $P^s = (p_{x,y}^s)$ is strongly connected. In general, we can see that the linear programming approach can be used for an arbitrary Markov decision problem where an arbitrary stationary strategy generates a unichain. We call such Markov decision processes *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 $P^s = (p_{x,y}^s)$ contains at least $[(|X| + 1)/2] + 1$ nonzero elements then the corresponding graph $G_p^s = \left(X, E_p^s\right)$ contains a unique strongly connected component that can be reached from every $x \in X$ [19], i.e., in this case the matrix P^s corresponds to a Markov unichain.

Let $s : X \to A$ be an arbitrary strategy ($s \in S$) for a 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

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

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

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

Lemma 2.25 *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*

$$\psi(s, q) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} s_{x,a}\, q_x \tag{2.64}$$

subject to

$$\begin{cases} \displaystyle\sum_{x \in X} \sum_{a \in A(x)} p^a_{x,y} s_{x,a} q_x = q_y, \quad \forall y \in X; \\ \displaystyle\sum_{x \in X} q_x = 1; \\ \displaystyle\sum_{a \in A(x)} s_{x,a} = 1, \quad \forall x \in X; \\ \qquad s_{x,a} \in \{0, 1\}, \ \ \forall x \in X, \ a \in A(x); \quad q_x \geq 0, \ \ \forall x \in X, \end{cases} \tag{2.65}$$

where

$$\mu_{x,a} = \sum_{y \in X} c^a_{x,y}\, p^a_{x,y}$$

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 (2.65) has a unique solution with respect to 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 of the objective function (2.64) for this solution expresses the average cost per transition for an arbitrary fixed starting state. Therefore, for a 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 (2.64), (2.65) for the perfect Markov decision process then we obtain the optimal stationary strategy s^*. □

Remark 2.26 For a perfect Markov decision processes the objective function $\psi(s, q)$ on the set of feasible solutions depends 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 (2.65) hold 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 (2.65) are essential for non perfect Markov processes.

Based on Lemma 2.25 we can prove the following result.

Theorem 2.27 *Let $\alpha^*_{x,a}$ $(x \in X, \ a \in A(x)), \quad q^*_x$ $(x \in X)$ be a basic optimal solution of the following linear programming problem:*
Minimize

$$\overline{\psi}(\alpha, q) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha_{x,a} \tag{2.66}$$

subject to

$$\begin{cases} \displaystyle\sum_{x \in X} \sum_{a \in A(x)} p^a_{x,y} \, \alpha_{x,a} = q_y, \quad \forall y \in X; \\ \displaystyle\sum_{x \in X} q_x = 1; \\ \displaystyle\sum_{a \in A(x)} \alpha_{x,a} = q_x, \quad \forall x \in X; \\ \alpha_{x,a} \geq 0, \ \ \forall x \in X, \ a \in A(x); \ q_x \geq 0, \ \forall x \in X, \end{cases} \tag{2.67}$$

where

$$\mu_{x,a} = \sum_{y \in X} c^a_{x,y} \, p^a_{x,y} \ \ for \ \ x \in X.$$

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

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

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 $\overline{\psi}(\alpha^*, q^*)$, i.e.,

$$f_x(s^*) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha^*_{x,a}$$

for every $x \in X$.

Proof The proof of this theorem is similar to the proof of Theorem 2.3. Applying Lemma 2.25 we obtain that the bilinear programming problem (2.64), (2.65) with boolean variables $s_{x,a}$ for $x \in X,\ a \in A(x)$ can be reduced to the linear programming problem (2.66), (2.67). We observe that the restriction $s_{x,a} \in \{0, 1\}$ in the problem (2.64), (2.65) can be replaced 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

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

can be changed by the restrictions

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

because the condition $q_x > 0, \quad \forall x \in X$ for the perfect Markov process holds. This means that the system (2.65) in the problem (2.64), (2.65) can be replaced by the following system

$$\begin{cases} \displaystyle\sum_{x \in X} \sum_{a \in A(x)} p^a_{x,y}\, s_{x,a}\, q_x = q_y, \quad \forall y \in X; \\ \displaystyle\sum_{x \in X} q_x = 1; \\ \displaystyle\sum_{a \in A(x)} s_{x,a}\, q_x = q_x, \quad \forall x \in X; \\ \qquad s_{x,a} \geq 0, \quad \forall x \in X,\ a \in A(x); \quad q_x \geq 0, \quad \forall x \in X. \end{cases} \tag{2.68}$$

In such a way we may conclude that problem (2.64), (2.65) and problem (2.64), (2.68) have the same optimal solutions. Taking into account that for the perfect network we have $q_x > 0, \quad \forall x \in X$ then in problem (2.64), (2.68) we can introduce the notations $\alpha_{x,a} = s_{x,a}\, q_x$ for $x \in X,\ a \in A(x)$, i.e., we obtain the problem (2.66), (2.67). 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. □

It is easy to observe that q_x in the system (2.67) 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 2.27 can be formulated in the following way.

Theorem 2.28 *Let $\alpha^*_{x,a}$ $(x \in X,\ a \in A(x))$, be a basic optimal solution of the following linear programming problem:*

Minimize

$$\overline{\psi}(\alpha) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha_{x,a} \tag{2.69}$$

subject to

$$\begin{cases} \sum\limits_{a \in A(y)} \alpha_{y,a} - \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, \alpha_{x,a} = 0, \quad \forall\, y \in X; \\ \sum\limits_{x \in X} \sum\limits_{a \in A(x)} \alpha_{x,a} = 1; \\ \alpha_{x,a} \geq 0, \quad \forall x \in X,\ a \in A(x). \end{cases} \tag{2.70}$$

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

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

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 $\overline{\psi}(\alpha^*, q^*)$, 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, based on theorems proven above the optimal stationary strategy for the Markov decision problem can be found using the following algorithm.

Algorithm 2.29 Determining the Optimal Stationary Strategies for the Perfect Markov Decision Problem

(1) Formulate the linear programming problem (2.66), (2.67) and find a basic optimal solution $\alpha^*_{x,y}$, q^*_x, q^*_z of this problem;
(2) Fix $s^*_{x,a} = 1$ for (x, a) that corresponds to the basic components of the optimal solution and set $s^*_{x,a} = 0$ for the remaining components.

Example Consider the Markov decision problem with an average cost criterion from Sect. 2.3.2. The corresponding multigraph of the Markov decision process is represented in Fig. 2.9.

The optimal stationary strategy s^* of this problem can be found by solving the linear programming problem (2.66), (2.67), i.e.:
Minimize

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

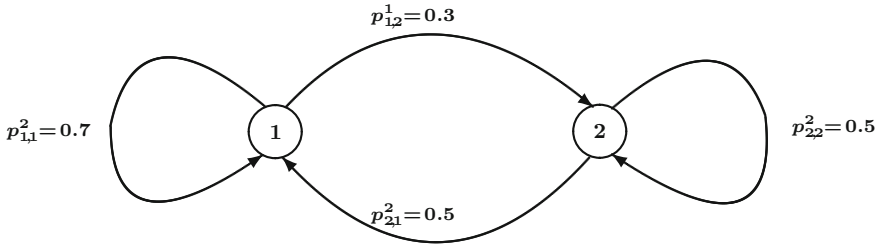

Fig. 2.12 The graph induced by the optimal strategy in Markov decision problem

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

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

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

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, the optimal stationary strategy s^* determines the Markov process with the following probability and cost matrices

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

The graph of transition probabilities of this Markov process is represented in Fig. 2.12.

The result described above shows that the Markov decision problem with an average cost criterion can be transformed into a stochastic optimal control problem on the auxiliary network $(G', X_C, X_N, p', c', x_0)$. This means that the linear programming algorithm proposed in the previous sections can be developed and specified for Markov decision problems with an average and discounted costs optimization criteria.

2.3.4 A Dual Linear Programming Model for an Average Markov Decision Problem

Consider the linear programming problem (2.69), (2.70) for an arbitrary unichain Markov decision process. As we have shown the solution of this problem always exists. If we dualize (2.69), (2.70) then we obtain the following problem:
Maximize

$$\psi'(\varepsilon, \omega) = \omega \tag{2.71}$$

subject to

$$\varepsilon_x - \sum_{y \in X} p^a_{x,y} \varepsilon_y + \omega \leq \mu_{x,a}, \quad \forall x \in X, \ \forall a \in A. \tag{2.72}$$

Based on duality theory of linear programming we obtain the following result.

Theorem 2.30 *The linear programming problem* (2.71), (2.72) *has solutions and an arbitrary optimal solution* ε^*, ω^* *of the problem possesses the following property: For each* $x \in X$ *there exists an action* $a^* \in A(x)$ *that satisfies the condition*

$$\min_{a \in A(x)} \left\{ \mu_{x,a^*} + \sum_{y \in X} p^{a^*}_{x,y} \varepsilon^*_y - \varepsilon^*_x - \omega^* \right\} = 0, \quad \forall x \in X. \tag{2.73}$$

The action a^* *in each state* $x \in X$ *determines the optimal stationary strategy* $s^*(x) = a^*$ *and* ω^* *is equal to the optimal value of the average cost in the Markov decision process.*

This theorem represents the optimization criterion for unichain Markov decision problems with average expected cost. Based on this criterion we can determine the optimal stationary strategies of the problem in the unichain case using the following algorithm.

Algorithm 2.31 Determining the Optimal Solution of a Unichain Markov Decision Problem Using a Dual Linear Programming Model

(1) Formulate the linear programming problem (2.71), (2.72) and find an optimal solution ε^*, ω^* of this problem;
(2) For each $x \in X$ fix $s^*(x) = a^*$, where $a^* \in A(x)$ satisfies condition (2.73).

2.3.5 Optimality Conditions for Multichain Decision Problems and a Linear Programming Approach

The optimality conditions for a *multichain Markov decision problem* with an average optimization cost criterion can be derived from the optimality conditions for an average multichain control problem if we take into account the mentioned relationship between Markov decision processes and stochastic control models. Based on Theorem 2.13 and the results from Sect. 2.3.2 we can formulate the following optimality principle for an average multichain decision problem.

Theorem 2.32 *Let a Markov decision process* (X, A, p, c) *be given. Then the system of equations*

$$\varepsilon_x + \omega_x = \min_{a \in A(x)} \left\{ \mu_{x,a} + \sum_{y \in X} p^a_{x,y} \varepsilon_y \right\}, \quad \forall x \in X; \tag{2.74}$$

has a solution under the set of solutions of the system of equations

$$\omega_x = \min_{a \in A(x)} \left\{ \sum_{y \in X} p^a_{x,y} \omega_x \right\}, \qquad \forall x \in X, \tag{2.75}$$

i.e., the system of equations (2.75) *has such a solution* $\omega^*_x,\ x \in X$ *for which there exists a solution* $\varepsilon^*_x,\ x \in X$ *of the system of equations*

$$\varepsilon_x + \omega^*_x = \min_{a \in A(x)} \left\{ \mu_{x,a} + \sum_{y \in X} p^a_{x,y} \varepsilon_y \right\}, \quad \forall x \in X. \tag{2.76}$$

The values ω^*_x *for* $x \in X$ *coincide with the optimal average costs* ω_x, $x \in X$ *for the Markov decision problem and an optimal stationary strategy*

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

for an average Markov decision problem can be found by fixing a map $s^*(x) = a \in A(x)$ *such that*

$$a \in \operatorname*{argmin}_{a \in A(x)} \left\{ \sum_{y \in X} p^a_{x,y} \omega^*_x \right\}$$

and

$$a \in \operatorname*{argmin}_{a \in A(x)} \left\{ \mu_{x,a} + \sum_{y \in X} p^a_{x,y} \varepsilon^*_y \right\}.$$

Note that the Lemmas 2.14, 2.20 are valid also for average Markov decision problems if the strategies s and s^* we treat as strategies $s : X \to A$; $s^* : X \to A$ for the Markov decision problem.

Then from these lemmas we obtain the proof of Theorem 2.32. The proof of this theorem also follows from the Theorems 2.12, 2.13 and the reduction procedure from Markov decision problem to stochastic control problem described in Sect. 2.3.2.

From Theorem 2.32 we can make the following conclusion. To determine a solution of the Markov decision problem it is necessary to determine ω_x for $x \in X$ that satisfies (2.75) and for which there exists ε_x for $x \in X$ that satisfies (2.74). This is equivalent with the problem of determining the "maximal" vector ω with the components ω_x for $x \in X$ that satisfies the conditions

$$\varepsilon_x + \omega_x \le \mu_{x,a} + \sum_{y \in X} p^a_{x,y} \varepsilon_y, \quad \forall x \in X, \ \ \forall a \in A(x);$$

$$\omega_x \le \sum_{y \in X} p^a_{x,y} \omega_x, \quad \forall x \in X, \ \ \forall a \in A(x).$$

Thus, we have to maximize a positive linear combination of components of ω under the restrictions given above, i.e., we obtain the following linear programming problem:
Maximize

$$\psi'(\varepsilon, \omega) = \sum_{x \in X} \theta_x \omega_x \tag{2.77}$$

subject to

$$\begin{cases} \varepsilon_x + \omega_x \le \mu_{x,a} + \sum\limits_{y \in X} p^a_{x,y} \varepsilon_y, & \forall x \in X, \ \ \forall a \in A(x); \\ \omega_x \le \sum\limits_{y \in X} p^a_{x,y} \omega_x, & \forall x \in X, \ \ \forall a \in A(x) \end{cases} \tag{2.78}$$

where $\theta > 0$, $\forall x \in X$ and $\sum_{x \in X} \theta_x = 1$.

From Theorem 2.32 we obtain the following result.

Corollary 2.33 *For an arbitrary strategy $s : X \to A$ the following system of linear equations*

$$\begin{cases} \varepsilon_x + \omega_x = \mu_{x,s(x)} + \sum\limits_{y \in X} p^{s(x)}_{x,y} \varepsilon_y, & \forall x \in X; \\ \omega_x = \sum\limits_{y \in X} p^{s(x)}_{x,y} \omega_x, & \forall x \in X \end{cases} \tag{2.79}$$

has a solution.

2.3.6 *Primal and Dual Linear Programming Models for a Multichain Markov Decision Problem*

We can regard the linear programming model (2.77), (2.78) as a dual model for a *primal multichain linear programming problem*. So, if we consider the dual model or (2.77), (2.78) then we obtain the following linear programming problem:
Minimize

$$\overline{\psi}(\alpha) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha_{x,a} \tag{2.80}$$

subject to

$$\begin{cases} \sum\limits_{a \in A(y)} \alpha_{y,a} - \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y} \alpha_{x,a} = 0, \ \ \forall y \in X; \\ \sum\limits_{a \in A(y)} \alpha_{y,a} + \sum\limits_{a \in A(y)} \beta_{y,a} - \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y} \beta_{x,a} = \theta_y, \ \forall y \in X; \\ \alpha_{x,a} \geq 0, \ \beta_{y,a} \geq 0, \ \ \forall x \in X, \ a \in A(x), \end{cases} \tag{2.81}$$

where $\theta > 0, \ \forall y \in X$ and $\sum_{y \in X} \theta_y 1$.
This problem generalizes the unichain linear programming problem (2.69), (2.70) from Sect. 2.3.3. In (2.81) the restrictions

$$\sum_{a \in A(y)} \alpha_{y,a} + \sum_{a \in A(y)} \beta_{y,a} - \sum_{x \in X} \sum_{a \in A(x)} p^a_{x,y} \beta_{x,a} = \theta_y, \ \ \forall y \in X \tag{2.82}$$

with the condition $\sum_{y \in X} \theta_y = 1$ generalize the constrain

$$\sum_{x \in X} \sum_{a \in A(y)} \alpha_{y,a} = 1 \tag{2.83}$$

in the unichain model. It is easy to check that by summing (2.82) over y, we obtain the equality (2.83).

2.4 Iterative Algorithms for Markov Decision Processes and Control Problems with an Average Cost Criterion

As we have shown the Markov decision problem and optimal control problems with average cost criterion can be solved using the linear programming approach. Here we show that these problems can be solved using iterative algorithms. These algorithms are based on the optimization criteria proved in previous sections. We can observe

that the optimization criterion for a stochastic control problem in the case $X_C = \emptyset$ leads to the equation which can be derived directly from formula (1.59).

Indeed, using formula (1.59) we can write the following two equivalent equations

$$\sigma(t) = t\omega + \varepsilon + \epsilon(t),$$
$$\sigma(t-1) = (t-1)\omega + \varepsilon + \epsilon(t-1),$$

where $\epsilon(t)$ and $\epsilon(t-1)$ tend to zero if t tends to infinity. If we introduce the expression of $\sigma(t)$ and $\sigma(t-1)$ in the recursive formula

$$\sigma(t) = \mu + P\sigma(t-1)$$

then we obtain

$$t\omega + \varepsilon + \epsilon(t) = \mu + P((t-1)\omega + \varepsilon + \epsilon(t-1).$$

Through rearrangement we get

$$\varepsilon + t\omega - (t-1)P\omega = \mu + P\varepsilon + P\epsilon(t-1) - \epsilon(t).$$

Here $\omega = P\omega$. In addition for a Markov unichain all components of the vector ω are the same, i.e., $\omega_1 = \omega_2 = \cdots = \omega_n = \omega$ (here $\omega_i = \omega_{x_i}$). So, if $t \to \infty$ then $\epsilon(t),\ \epsilon(t-1) \to 0$ and we obtain

$$\varepsilon_i + \omega = \mu_{x_i} + [P\varepsilon]_i, \quad i = 1, 2, \ldots, n. \tag{2.84}$$

This is the system of equations for a unichain Markov process. It is well known that in the case of unichain processes the rank of the matrix $(I - P)$ is equal to $n - 1$ (see [98]). Based on this fact in [98] it has been shown that the system of equations (2.84) has a unique solution once it is setting $\varepsilon_i = 0$ for some i. This means that two different vectors ε' and ε'' which represent the solutions of this equation differ only by some constant for each component. Therefore, the system of equations (2.84) allows us to determine the average cost per transition in unichain Markov processes with transition costs. The existence of the solution of this system of equations (2.84) also follows from Theorem 2.30.

The system of equations for the decision problem in the case of unichain processes is the following

$$\varepsilon_i + \omega = \min_{a \in A(x_i)} \left(\mu_{x_i,a} + [P^a\varepsilon]_i\right), \quad i = 1, 2, \ldots, n. \tag{2.85}$$

According to Theorem 2.30 the system of equations (2.85) has solutions. The solution of this system of equations and the optimal stationary strategy for unichain Markov decision problems can be found using the following iterative algorithm.

Algorithm 2.34 Determining the Solution of a Unichain Markov Decision Problem

Preliminary step (Step 0): Fix an arbitrary stationary strategy

$$s^0 : x_i \to a \in A(x_i) \ \text{ for } \ x_i \in X.$$

General step (Step k, $k > 0$): Calculate

$$\mu_{x_i,a^{k-1}} = \sum_{y \in X(x_i)} p_{x_i,y}^{s^{k-1}(x_i)} c_{x_i,y}^{s^{k-1}(x_i)}$$

for every $x_i \in X$. Then solve the system of linear equations

$$\varepsilon_i^{k-1} + \omega^{k-1} = \mu_{x_i,s^{k-1}(x_i)} + [P^{s^{k-1}} \varepsilon^{k-1}]_i, \quad i = 1, 2, \ldots, n,$$

$$\varepsilon_n^{k-1} = 0,$$

and find $\varepsilon_1^{k-1}, \varepsilon_2^{k-1}, \ldots, \varepsilon_{n-1}^{k-1}$ and ω^{k-1}. After that determine a new strategy

$$s^k : x_i \to a \in A(x_i) \ \text{ for } \ x_i \in X,$$

where

$$s^k(x_i) = \operatorname*{argmin}_{a \in A(x_i)} \left(\mu_{x_i,a} + [P^a \varepsilon^{k-1}]_i\right), \quad i = 1, 2, \ldots, n.$$

Check if the following condition holds

$$s^k(x_i) = s^{k-1}(x_i), \quad \forall x_i \in X. \tag{2.86}$$

If the condition (2.86) holds then fix

$$s^* = s^k, \quad \omega^* = \omega^k$$

as the optimal solution of the problem; otherwise go to the next step $k + 1$.

The correctness and the convergence of this algorithm follow from the results described above and the results from [115, 118–120, 140].

The algorithm described above can be specified for determining the optimal stationary strategies in the stochastic control problem with an average cost optimization criterion.

Algorithm 2.35 Determining the Solution for a Stochastic Control Problem

Let the average control problem on a perfect network determined by the graph of state's transition $G = (X, E)$ with the set of controllable states X_C, the set of

uncontrollable states X_N, the probability function $p : E_N \to [0, 1]$ which satisfies the condition from Sect. 2.3 and the cost function $c : E \to \mathbb{R}$ be given.

Preliminary step (Step 0): Fix an arbitrary stationary strategy

$$s^0 : x_i \to x_j \in X(x_i) \text{ for } x_i \in X_C.$$

General step (Step k, $k > 0$): Determine the probability matrix $P^{s^{k-1}} = \left(p^{s^{k-1}}_{x_i,x_j}\right)$, where

$$p^{s^{k-1}}_{x_i,x_j} = \begin{cases} p_{x_i,x_j}, & \text{if } x_i \in X_N \text{ and } (x_i, x_j) \in E_2; \\ 1, & \text{if } x_i \in X_C \text{ and } x_j = s^{k-1}(x_i); \\ 0, & \text{if } x_i \in X_C \text{ and } x_j \neq s^{k-1}(x_i). \end{cases}$$

Then calculate

$$\mu_{x_i,s^{k-1}(x_i)} = \sum_{x_j \in X(x_i)} p^{s^{k-1}(x_i)}_{x_i,x_j} c_{x_i,x_j}$$

for every $x_i \in X$. After that solve the system of linear equations

$$\varepsilon^{k-1}_i + \omega^{k-1} = \mu_{x_i,s^{k-1}(x_i)} + [P^{s^{k-1}} \varepsilon^{k-1}]_i, \quad i = 1, 2, \ldots, n,$$

$$\varepsilon^{k-1}_n = 0,$$

and find $\varepsilon^{k-1}_1, \varepsilon^{k-1}_2, \ldots, \varepsilon^{k-1}_{n-1}$ and ω^{k-1}. Then determine a new strategy

$$s^k : x_i \to x_j \in X(x_i) \quad \text{for} \quad x_i \in X_C,$$

where

$$s^k(x_i) = \operatorname*{argmin}_{x_j \in X(x_i)} \left(c_{x_i,x_j} + \varepsilon^{k-1}_j\right), \quad \forall x_i \in X_C.$$

Check if the following condition holds

$$s^k(x_i) = s^{k-1}(x_i), \quad \forall x_i \in X_C. \tag{2.87}$$

If the condition (2.87) holds then fix

$$s^* = s^k, \quad \omega^* = \omega^k$$

as the optimal solution of the problem; otherwise go to the next step $k + 1$.

In the case $X_N = \emptyset$ this algorithm is transformed into the algorithm for solving a deterministic control problem. In this case the algorithm correctly finds the solution of

the problem if each stationary strategy in G generates a subgraph G_s which contains a unique directed cycle.

The algorithms described above determine the optimal stationary strategies for a Markov decision problem and a stochastic optimal control problem if an arbitrary strategy in these problems generates a unichain process.

For the multichain case of the problem the algorithm uses the multichain bias equations (2.74)–(2.76).

Algorithm 2.36 Determining the Solution of a Multichain Markov Decision Problem

Preliminary step (Step 0): Fix an arbitrary stationary strategy

$$s^0 : x_i \to a \in A(x_i) \ \text{ for } \ x_i \in X.$$

General step (Step k, $k \geq 1$): Determine the matrix $P^{s^{k-1}}$ and $\mu^{s^{k-1}}$ that corresponds to the strategy s^{k-1}. Find $\omega^{s^{k-1}}$ and $\varepsilon^{s^{k-1}}$ which satisfy the conditions

$$\begin{cases} (P^{s^{k-1}} - I)\omega^{s^{k-1}} = 0; \\ \mu^{s^{k-1}} + (P^{s^{k-1}} - I)\varepsilon^{s^{k-1}} - \omega^{s^{k-1}} = 0. \end{cases}$$

Then find a strategy s^k such that

$$s^k \in \operatorname*{argmin}_s \left\{ P^s \omega^{s^{k-1}} \right\}$$

and set $s^k = s^{k-1}$ if

$$s^{k-1} \in \operatorname*{argmin}_s \left\{ P^s \omega^{s^{k-1}} \right\}.$$

After that check if $s^k = s^{k-1}$? If $s^k = s^{k-1}$ then go to next step $k+1$; otherwise choose the strategy s^k such that

$$s^k \in \operatorname*{argmin}_s \left\{ \mu^s + P^s \varepsilon^{s^{k-1}} \right\}$$

and set $s^k = s^{k-1}$ if

$$s^{k-1} \in \operatorname*{argmin}_s \left\{ \mu^s + P^s \varepsilon^{s^{k-1}} \right\}.$$

After that check if $s^k = s^{k-1}$? If $s^k = s^{k-1}$ then STOP and set $s^* = s^{k-1}$; otherwise go to the next step $k+1$.

The convergence of the algorithms based on iterative procedures are proved in [115, 121–123]. In a similar way as for the unichain case of the problem the algorithm

described above can be specified for a multichain stochastic control problem on networks. The computational complexity of the Markov decision problems in the general case is studied in [106].

2.5 A Discounted Stochastic Control Problem and Algorithms for Determining the Optimal Strategies on Networks

Now we consider the infinite horizon discounted stochastic control problem. Following the concept from the previous sections we formulate the discounted stochastic control problem on networks and describe algorithms for determining the optimal stationary strategies using a linear programming approach. Then we extend this approach for Markov decision problems with an expected total discounted cost optimization criterion.

2.5.1 Problem Formulation

Let a time-discrete system $\mathbb{L}$ with finite set of states X be given and assume that the dynamics of the system is described by a directed graph of states' transitions $G = (X, E)$ with the vertex set X and edge set E. Thus, an arbitrary directed edge $e = (x, y) \in E$ 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 $t = 0, 1, 2, \ldots$. On an edge set E a cost function $c : E \to \mathbb{R}$ is defined that indicates a cost c_e to each directed edge $e = (x, y) \in E$ if the system makes a transition from the state $x = x(t)$ to the state $y = x(t + 1)$ for every $t = 0, 1, 2, \ldots$. We define the stationary control for the system $\mathbb{L}$ in G as a map

$$s : x \to y \in X(x) \quad \text{for} \quad x \in X,$$

where $X(x) = \{y \in X \mid (x, y) \in E\}$.

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 expected total cost $\sigma_{x_0}^{\gamma}(s)$ is defined as follows

$$\sigma_{x_0}^{\gamma}(s) = \sum_{t=0}^{\infty} \gamma^t c_{e_t^s},$$

where $\gamma, \quad 0 < \gamma < 1$, is a given discount factor.

Based on the results from [47, 114] it is easy to show that for an arbitrary stationary strategy s there exists $\sigma^{\gamma}_{x_0}(s)$. If we denote by $\sigma^{\gamma}(s)$ the column vector with components $\sigma^{\gamma}_{x}(s)$ for $x \in X$ then $\sigma^{\gamma}_{x_0}(s)$ can be found by solving the system of linear equations

$$(I - \gamma P^s)\sigma^{\gamma}(s) = c^s, \tag{2.88}$$

where c^s is the vector with corresponding components $c_{x,s(x)}$ for $x \in X$, I is the identity matrix and P^s the matrix with elements $p^s_{x,y}$ for $x, y \in X$ defined as follows:

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

It is well known that for $0 < \gamma < 1$ the rank of the matrix $I - \gamma P^s$ is equal to $|X|$ and the system (2.88) has solutions for arbitrary c^s (see [114, 140]). Thus, we can determine $\sigma^{\gamma}_{x_0}(s^*)$ for an arbitrary starting state x_0.

In the considered deterministic discounted control problem on G we are seeking for a stationary control s^* such that

$$\sigma^{\gamma}_{x_0}(s^*) = \min_{s} \sigma^{\gamma}_{x_0}(s).$$

We formulate and study this problem in a more general case considering its *stochastic version*. 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 $X = X_C \cup X_N$, $X_C \cap X_N = \emptyset$, where X_C represents the set of states in which the decision maker is able to control the dynamical system and where X_N represents the set of states in which the dynamical system makes transitions to the next state in a random way. This means that for every $x \in X$ on the set of feasible transitions $E(x)$ the distribution function $p : E(x) \to \mathbb{R}$ is defined such that $\sum_{e \in E(x)} p_e = 1$, $p_e \geq 0$, $\forall e \in E(x)$ and the transitions from the states $x \in X_N$ to the next states are made randomly according to these distribution functions. Here, in a similar way as for the deterministic problem we assume that to each directed edge $e = (x, y) \in E$ a cost c_e of system's transition from the state $x = x(t)$ to the state $y = x(t + 1)$ for $t = 0, 1, 2, \ldots$ is associated. In addition we assume that the discount factor γ, $0 < \gamma < 1$, and the starting state x_0 are given. We define a stationary control on G as a map

$$s : x \to y \in X(x) \quad \text{for} \quad x \in X_C.$$

Let s be an arbitrary stationary strategy. We define the graph $G_s = (X, E_s \cup E_N)$, where $E_s = \{e = (x, y) \in E \,|\, x \in X_C, y = s(x)\}$, $E_N = \{e = (x, y) \,|\, x \in X_N, y \in X\}$. This graph corresponds to a Markov process with the probability matrix $P^s = (p^s_{x,y})$, where

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

For this Markov process with associated costs c_e, $e \in E$ we can define the expected total discounted cost $\sigma^\gamma_{x_0}(s)$ as we have introduced in Chap. 1. We consider the problem of determining the strategy s^* for which

$$\sigma^\gamma_{x_0}(s^*) = \min_s \sigma^\gamma_{x_0}(s).$$

Without loss of generality we may consider that G has the property that an arbitrary vertex in G is reachable from x_0; otherwise we can delete all vertices that could not be reached from x_0.

2.5.2 A Linear Programming Approach for a Discounted Control Problem on Networks

We develop a linear programming approach for the discounted stochastic control problem on the network (G, X, E, c, p, x_0) with a given discount factor γ using the same logical scheme as in Sect. 2.2. We identify an arbitrary stationary strategy s in G with the set of boolean variables $s_{x,y}$ for $x \in X_C$ and $y \in X(x)$, where

$$s_{x,y} = \begin{cases} 1, & \text{if } y = s(x); \\ 0, & \text{if } y \neq s(x). \end{cases} \tag{2.89}$$

In the following we will simplify the notations and instead σ^γ_x we shall use σ_x.

Lemma 2.37 *For a fixed strategy s the values σ^γ_x, $x \in X$ determine the unique optimal basic solution of the following linear programming problem:*
Maximize

$$\varphi^s_{x_0}(\sigma) = \sigma_{x_0} \tag{2.90}$$

subject to

$$\begin{cases} \sigma_x - \gamma \displaystyle\sum_{y \in X(x)} s_{x,y}\, \sigma_y \;\leq\; \sum_{y \in X(x)} c_{x,y}\, s_{x,y}\,, & \forall x \in X_C; \\ \sigma_x - \gamma \displaystyle\sum_{y \in X(x)} p_{x,y}\, \sigma_y \;\leq\; \mu_x, & \forall x \in X_N; \end{cases} \tag{2.91}$$

where

$$\mu_x = \sum_{y \in X(x)} c_{x,y}\, p_{x,y}, \quad \forall x \in X_N$$

and

$$\sum_{y \in X(x)} s_{x,y} = 1, \ \ \forall x \in X_C; \ \ s_{x,y} \in \{0, 1\}, \ \ \forall x \in X_C, \ y \in X.$$

Proof If for a fixed strategy s we treat the values $s_{x,y}$ as the transition probabilities from the states $x \in X$ to the states $y \in X$ then the condition (2.88) in the extended form can be written as follows:

$$\begin{cases} \sigma_x - \gamma \sum\limits_{y \in X(x)} s_{x,y}\, \sigma_y = \sum\limits_{y \in X(x)} c_{x,y}\, s_{x,y}, & \forall x \in X_C; \\ \sigma_x - \gamma \sum\limits_{y \in X(x)} p_{x,y}\, \sigma_y = \mu_x, & \forall x \in X_N. \end{cases} \tag{2.92}$$

This system determines uniquely the values σ_x for $x \in X$. Therefore, for fixed s the linear programming problem (2.90), (2.92) has a solution and the optimal value of the objective function is equal to σ_{x_0}.

It is evident that if in this system we change the costs $c_{x,y}$ by new costs $c'_{x,y}$ such that $c'_{x,y} \le c_{x,y}$ then we obtain a new linear programming problem for which the corresponding optimal value σ'_{x_0} of the objective function is less or equal to σ_{x_0}. Thus, if we change the system of linear equations (2.92) by the following system of linear inequalities

$$\begin{cases} \sigma_x - \gamma \sum\limits_{y \in X(x)} s_{x,y}\, \sigma_y \le \sum\limits_{y \in X(x)} c_{x,y}\, s_{x,y}, & \forall x \in X_C; \\ \sigma_x - \gamma \sum\limits_{y \in X(x)} p_{x,y}\, \sigma_y \le \mu_x, & \forall x \in X_N, \end{cases} \tag{2.93}$$

then for a fixed strategy s we obtain a new linear programming problem (2.90), (2.93) with the optimal solution σ_{x_0}. So, the lemma holds. □

Now we consider the optimization problem (2.90), (2.93) in the case if $s_{x,y}$ are arbitrary boolean variables and correspond to the possible stationary strategies. So, if we add the condition $s_{x,y} \in \{0, 1\}, \ \ \forall x \in X_C, \ y \in X$ to (2.93) then we obtain the mixed integer bilinear programming problem in which we have to maximize with respect to σ_x and minimize with respect to s.

Based on Lemma 2.37 we can prove the following result.

Theorem 2.38 *Let $\alpha^*_{x,y}$ $(x \in X_C,\ y \in X)$, β^*_x $(x \in X)$ be an optimal solution of the following linear programming problem:*
Minimize

$$\phi_{x_0}(\alpha, \beta) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y}\, \alpha_{x,y} + \sum_{x \in X_N} \mu_x \beta_x \tag{2.94}$$

subject to

$$\begin{cases} \beta_y - \gamma \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - \gamma \sum\limits_{x \in X_N^-(y)} p_{x,y}\, \beta_x = 1, \quad y = x_0; \\ \beta_y - \gamma \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - \gamma \sum\limits_{x \in X_N^-(y)} p_{x,y}\, \beta_x = 0, \quad \forall y \in X \backslash \{x_0\}; \\ \sum\limits_{y \in X(x)} \alpha_{x,y} = \beta_x, \quad \forall x \in X_C; \\ \beta_x \geq 0,\ \forall x \in X; \quad \alpha_{x,y} \geq 0, \quad \forall x \in X_C,\ y \in X(x), \end{cases} \tag{2.95}$$

where

$$\mu_x = \sum_{y \in X(x)} c_{x,y}\, p_{x,y}, \quad \forall x \in X_N.$$

Then

$$\frac{\alpha^*_{x,y}}{\beta^*_x} \in \{0, 1\},\ \forall y \in X(x), \forall y \in X^*_C,$$

where $X_C^+ = \{x \in X_C \mid \beta_x > 0\}$ and an optimal stationary strategy for the discounted stochastic control problem on the network can be found as follows:

- *if $x \in X_C^+$ then fix*

$$s^*_{x,y} = \frac{\alpha^*_{x,y}}{\beta^*_x}, \forall x \in X(x);$$

- *if $x \in X \backslash X_C^+$ then fix an arbitrary $s_{x,y} \in \{0, 1\}$ for every $y \in X(x)$ such that*

$$\sum_{y \in X(x)} s_{x,y} = 1.$$

Proof According to Lemma 2.37 for a fixed strategy s the values $\sigma_x,\ x \in X$ can be found by solving the linear programming problem (2.90), (2.91).

Considering the dual problem for (2.90), (2.91) with respect to σ_x for a fixed strategy s we obtain the following optimization problem:
Minimize

$$\phi^s_{x_0}(\beta) = \sum_{x\in X_C} \sum_{y\in X(x)} c_{x,y}\, s_{x,y}\, \beta_x + \sum_{x\in X_N} \mu_x \beta_x \tag{2.96}$$

subject to

$$\begin{cases} \beta_y - \gamma \displaystyle\sum_{x\in X_C^-(y)} s_{x,y}\,\beta_x - \gamma \sum_{x\in X_N^-(y)} p_{x,y}\,\beta_x = 1, \quad y = x_0; \\ \beta_y - \gamma \displaystyle\sum_{x\in X_C^-(y)} s_{x,y}\,\beta_x - \gamma \sum_{x\in X_N^-(y)} p_{x,y}\,\beta_x = 0, \quad \forall y \in X\backslash\{x_0\}; \\ \beta_x \geq 0, \quad \forall x \in X. \end{cases} \tag{2.97}$$

In this system $s_{x,y}$ for $x \in X_C,\ y \in X(x)$ the following condition is satisfied:

$$\begin{cases} \displaystyle\sum_{y\in X(x)} s_{x,y} = 1, \quad \forall x \in X_C; \\ s_{x,y} \geq 0, \quad \forall x \in X,\ y \in X(x). \end{cases} \tag{2.98}$$

Then, an optimal strategy s^* of the control problem on G corresponds to an extreme point of the set of solutions of system (2.98). It is easy to observe that system (2.97) is consistent for an arbitrary feasible solution of system (2.97), and therefore, an optimal stationary strategy s^* can be determined by minimizing (2.96) with respect to $s_{x,y}$ and β_x subject to (2.97), (2.98). Thus, if we add condition (2.98) to condition (2.97) (and after that we minimize (2.96) with respect to $s_{x,y}$ and β_x), then we obtain the following nonlinear programming problem:
Minimize

$$\phi_{x_0}(s,\beta) = \sum_{x\in X_C} \sum_{y\in X(x)} c_{x,y} s_{x,y} \beta_x + \sum_{x\in X_N} \mu_x \beta_x \tag{2.99}$$

subject to

$$\begin{cases} \beta_y - \gamma \displaystyle\sum_{x\in X_C^-(y)} s_{x,y}\,\beta_x - \gamma \sum_{x\in X_N^-(y)} p_{x,y}\,\beta_x = 1, \quad y = x_0; \\ \beta_y - \gamma \displaystyle\sum_{x\in X_C^-(y)} s_{x,y}\,\beta_x - \gamma \sum_{x\in X_N^-(y)} p_{x,y}\,\beta_x = 0, \quad \forall y \in X\backslash\{x_0\}; \\ \displaystyle\sum_{y\in X(x)} s_{x,y} = 1, \quad \forall x \in X_C; \\ \beta_x \geq 0, \quad \forall x \in X; \quad s_{x,y} \geq 0, \quad \forall x \in X,\ y \in X(x). \end{cases} \tag{2.100}$$

This is a bilinear programming problem however it can be easily reduced to a linear programming problem (2.96), (2.97) using the following elementary transformations:

We change in (2.100) the restrictions $\sum_{y\in X(x)} s_{x,y} = 1,\ \forall x \in X_C$ by $\sum_{y\in X(x)} s_{x,y}\,\beta_y = \beta_y,\ \forall x \in X_C$ and then we introduce the notations

$$\alpha_{x,y} = s_{x,y}\,\beta_y,\quad \forall x \in X,\ y \in X(x). \tag{2.101}$$

It is easy to observe that if $\alpha^*_{x,y}\ (x \in X_C,\ y \in X),\quad \beta^*_y\ (y \in X)$ is a basic optimal solution of problem (2.94), (2.95) then for each $x \in X_C^+$ among $\alpha^*_{x,y},\ y \in X(x)$ only one it is different from zero and it is equal to β^*_x. Moreover, if $\alpha^*_{x,y}\ (x \in X_C,\ y \in X),\quad \beta^*_y\ (y \in X)$ is a basic optimal solution of the problem (2.94), (2.95) then $\alpha^*_{x,y} = 0,\ \forall x \in X\backslash X_C^+, \forall y \in X)$ and therefore for $x \in X\backslash X_C^+,\ y \in X(x)$ in the optimal solution of problem (2.99), (2.100) we can fix arbitrary $s^*_{x,y} \in \{0, 1\}$ for $y \in X(x)$ such that $\sum_{y\in X(x)} s_{x,y} = 1$. So, we can determine the optimal stationary strategy for the control problem on network according to the rule formulated in the theorem. □

Note that in the considered control problem with fixed stating state x_0 the vertices $x \in X_c^+$ of the graph G correspond to the states in which the decision person makes the optimal control. The vertices $x \in X\backslash X_C^+$ of graph G correspond to the states of the dynamical system that couldn't be reached in the process of the optimal control made by the decision person. Therefore for the optimal solution of the control problem on G with fixed starting state x_0 we can set $s^*_{x,y} = 0, \forall x \in X\backslash X_C^+, \forall y \in X(x)$. This does not affect the sense of the control problem on networks.

From Theorem 2.38 in the case $X = X_C$ (i.e. $X_N = \emptyset$) we obtain conditions for determining the optimal stationary strategies of the deterministic discounted control problem.

Corollary 2.39 *Let $X = X_C$ and $\alpha^*_{x,y}\ (x \in X,\ y \in X)$, $\beta^*_x\ (x \in X)$ be an optimal solution of the following linear programming problem:*
Minimize

$$\phi_{x_0}(\alpha, \beta) = \sum_{x\in X} \sum_{y\in X(x)} c_{x,y}\,\alpha_{x,y} \tag{2.102}$$

subject to

$$\begin{cases} \beta_y - \gamma \displaystyle\sum_{x\in X^-(y)} \alpha_{x,y} = 1, \quad y = x_0; \\ \beta_y - \gamma \displaystyle\sum_{x\in X^-(y)} \alpha_{x,y} = 0, \quad \forall y \in X\backslash\{x_0\}; \\ \displaystyle\sum_{y\in X(x)} \alpha_{x,y} = \beta_x, \quad \forall x \in X; \\ \beta_x \geq 0,\ \forall x \in X; \quad \alpha_{x,y} \geq 0, \quad \forall x \in X,\ y \in X(x). \end{cases} \tag{2.103}$$

Then the optimal stationary strategy s^* of the discounted stochastic control problem on network can be found by fixing

$$s^*_{x,y} = \frac{\alpha^*_{x,y}}{\beta^*_x}, \quad \forall x \in X^+_C, \ y \in X(x)$$

and arbitrary $s^*_{x,y} \in \{0, 1\}$ for $x \in X \backslash X^+_C$ such that $\sum_{y \in X(x)} s_{x,y} = 1$;

Based on Theorem 2.38 we can propose the following algorithm for determining the optimal solution of the discounted control problem on the network.

Algorithm 2.40 Determining the Optimal Stationary Strategy for the Discounted Stochastic Control Problem

(1) Formulate the linear programming problem (2.96), (2.97);
(2) Determine an optimal solution $\alpha^*_{x,y}$ $(x \in X_C, \ y \in X)$, β^*_y $(y \in X)$ of the problem (2.96), (2.97) and fix

$$s^*_{x,y} = \frac{\alpha^*_{x,y}}{\beta^*_x}, \quad \forall x \in X_C, \ y \in X(x)$$

and an arbitrary $s^*_{x,y} \in \{0, 1\}$ for every $x \in X \backslash X^+_C$ such that $\sum_{y \in X(x)} s_{x,y} = 1$.

The results described above allow us to determine the stationary strategy for the problem with a fixed starting state x_0. In the general case, if it is necessary to find the optimal stationary strategy for an arbitrary starting state $x \in X$ then we can use the following results.

Lemma 2.41 *For a fixed strategy s the values σ^γ_x, $x \in X$ determine the unique optimal basic solution of the following linear programming problem:*
Maximize

$$\varphi^s(\sigma) = \sum_{x \in X} \sigma_x \tag{2.104}$$

subject to (2.91).

The proof of this lemma is identical to the proof of Lemma 2.37. Based on this lemma we can prove the following theorem.

Theorem 2.42 *Let $\alpha^*_{x,y}$ $(x \in X_C, \ y \in X)$, β^*_x $(x \in X)$ be a basic optimal solution of the following linear programming problem:*
Minimize

$$\phi(\alpha, \beta) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y} \, \alpha_{x,y} + \sum_{x \in X_N} \mu_x \beta_x \tag{2.105}$$

subject to

$$\begin{cases} \beta_y - \gamma \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - \gamma \sum\limits_{x \in X_N^-(y)} p_{x,y}\, \beta_x = 1, \quad \forall y \in X; \\ \sum\limits_{y \in X(x)} \alpha_{x,y} = \beta_x, \quad \forall x \in X_C; \\ \beta_x \geq 0, \quad \forall x \in X; \quad \alpha_{x,y} \geq 0, \quad \forall x \in X,\ y \in X(x). \end{cases} \tag{2.106}$$

If in the graph $G = (X, E)$ each vertex $x \in X$ contains at least one leaving directed edge then $\beta_x^* > 0$, $\forall x \in X_C$ and

$$\frac{\alpha_{x,y}^*}{\beta_x^*} \in \{0, 1\}, \quad \forall x \in X_C,\ y \in X(x).$$

The optimal stationary strategy s^* of the discounted stochastic control problem on the network can be found by fixing

$$s_{x,y}^* = \frac{\alpha_{x,y}^*}{\beta_x^*}, \quad \forall x \in X_C,\ y \in X(x).$$

The proof of this theorem is similar to the proof of Theorem 2.38 and the solution of the problem can be found by using the linear programming problem (2.105), (2.106). Based on this theorem we determine the optimal stationary strategies for an arbitrary starting state $x \in X$.

In the problem (2.105), (2.105) we can eliminate β_y from those restrictions that correspond to vertices $y \in X_C$ if we take into account the relation $\beta_y = \sum_{x \in X(y)} \alpha_{y,x}$ for $y \in X_C$. After that from Theorem 2.42 we obtain the following corollary:

Corollary 2.43 *Let $\alpha_{x,y}^*$ $(x \in X_C,\ y \in X)$, β_x^* $(x \in X)$ be a basic optimal solution of the following linear programming problem:*
Minimize

$$\phi(\alpha, \beta) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y}\, \alpha_{x,y} + \sum_{x \in X_N} \mu_x \beta_x \tag{2.107}$$

subject to

$$\begin{cases} \sum\limits_{x \in X(y)} \alpha_{y,x} - \gamma \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - \gamma \sum\limits_{x \in X_N^-(y)} p_{x,y}\, \beta_x = 1, \quad y \in X_C; \\ \beta_y - \gamma \sum\limits_{x \in X_C^-(y)} \alpha_{x,y} - \gamma \sum\limits_{x \in X_N^-(y)} p_{x,y}\, \beta_x = 1, y \in X_N; \\ \beta_x \geq 0, \ \forall x \in X_N; \quad \alpha_{x,y} \geq 0, \quad \forall x \in X_C,\ y \in X(x), \end{cases} \tag{2.108}$$

If in the graph $G = (X, E)$ each vertex $x \in X$ contains at least one leaving directed edge then $\sum_{y \in X(x)} \alpha^*_{x,y} > 0,\ \forall x \in X_C$ and

$$\frac{\alpha^*_{x,y}}{\sum_{y \in X} \alpha^*_{x,y}} \in \{0, 1\}, \quad \forall x \in X_C,\ y \in X(x).$$

The optimal stationary strategy s^* of the discounted stochastic control problem on the network can be found by fixing

$$s^*_{x,y} = \frac{\alpha^*_{x,y}}{\sum_{y \in X(x)} \alpha_{x,y}}, \quad \forall x \in X_C,\ y \in X(x).$$

2.5.3 Dual Linear Programming Models for a Discounted Control Problem

If we dualize the linear programming problem (2.94), (2.95) then on the basis of duality theory we obtain the following result.

Theorem 2.44 *Let w^*_x $(x \in X_C)$, σ^*_x $(x \in X)$ be the optimal solution of the linear programming problem:*
Maximize

$$\varphi_{x_0}(\sigma, w) = \sigma_{x_0} \tag{2.109}$$

subject to

$$\begin{cases} w_x - \gamma\sigma_y \le c_{x,y}, & \forall x \in X_C,\ y \in X(x); \\ -w_x + \sigma_x \le 0, & \forall x \in X_C; \\ \sigma_x - \gamma \sum\limits_{y \in X(x)} p_{x,y}\sigma_y \le \mu_x, & \forall x \in X_N. \end{cases} \tag{2.110}$$

Then $w^*_x = \sigma^*_x,\ \forall x \in X_C$ and $\sigma^*_{x_0}$ is the optimal discounted expected total cost for the problem on the network with the starting state x_0. An optimal stationary strategy can be found by fixing $s^* : X_C \to X$ such that $(x, s^*(x)) \in E^*(x),\ \forall x \in X_C$, where $E^*(x) = \{(x, y) \mid y \in X(x),\ \sigma^*_x - \gamma\sigma^*_y - c_{x,y} = 0\}$.

As a consequence from this theorem we obtain the following result.

Corollary 2.45 *For an arbitrary discounted control problem on the network (G, X_C, X_N, c, p) with a given discount factor γ there exist the values σ^*_x for $x \in X$ that satisfy the following conditions:*

(1) $\bar{c}_{x,y} = c_{x,y} + \gamma\sigma^*_y - \sigma^*_x \ge 0,\ \forall x \in X_C, y \in X(x)$;
(2) $\min\limits_{y \in X(x)} \{\bar{c}_{x,y}\} = 0,\ \forall x \in X_C$;

(3) $\overline{\mu}_x = \mu_x + \gamma \sum_{y \in X(z)} p_{x,y}\sigma^*_y - \sigma^*_x = 0, \ \forall x \in X_N.$

An arbitrary stationary strategy $s^* : X_C \to X$ *such that* $(x, s^*(x)) \in E^*(x)$, $\forall x \in X_C$, *where* $E^*(x) = \{(x, y) \mid y \in X(x), \overline{c}_{x,y} = 0\}$, *represents an optimal stationary strategy for the discounted control problem.*

In the case $X_N = \emptyset$, from this corollary we obtain the optimality condition for the deterministic discounted control problem. The conditions for determining the optimal strategy and the value of the optimal cost for the problem with $\gamma = 1$ in the case if this value exists can also be derived from Theorem 2.44 and Corollary 2.45.

The results formulated above can be extended to the problem of determining the optimal stationary strategy with an arbitrary starting state $x \in X$. For the problem (2.105), (2.106) we can construct the dual problem in a similar way and we then obtain the following result.

Theorem 2.46 *Let* σ^*_x, w^*_x $(x \in X)$ *be the optimal solution of the linear programming problem:*
Maximize

$$\varphi(\sigma, w) = \sum_{x \in X} \sigma_x \tag{2.111}$$

subject to (2.110). *Then* σ^*_x *for* $x \in X$ *represents the optimal discounted expected total costs for the problem on the network with starting states* $x \in X$. *An optimal stationary strategy can be found by fixing* $s^* : X_C \to X$ *such that* $(x, s^*(x)) \in E^*(x)$, $\forall x \in X_C$, *where* $E^*(x) = \{(x, y) \mid y \in X(x), \ \sigma^*_x - \gamma\sigma^*_y - c_{x,y} = 0\}$.

2.6 A Linear Programming Approach for a Discounted Markov Decision Problem

Consider a Markov decision process (X, A, p, c) with a finite set of states X, a finite set of actions A, the probability function $p : A \times X \times X \to [0, 1]$ that satisfies the condition $\sum_{y \in X} p^a_{x,y} = 1$, $\forall a \in A$ and the cost function $c : A \times X \times X \to \mathbb{R}$. In addition we assume that the discount factor γ, $0 \le \gamma < 1$, and the starting state x_0 are given.

Let us fix a stationary strategy

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

that induces a simple Markov process with a transition probability matrix $P^s = (p^s_{x,y})$ and a transition cost matrix $C^s = (c_{x,y})$. Then we can determine the discounted expected total costs $\sigma^\gamma_{x_0}(s)$ (in order to simplify the notation in the following we shall use $\sigma_{x_0}(s)$ instead of $\sigma^\gamma_{x_0}(s)$).

We consider the problem of determining the strategy s^* such that

$$\sigma_{x_0}(s^*) = \min_s \sigma_{x_0}(s).$$

In a similar way as for the control problem, here we identify an arbitrary strategy $s : X \to A$ with the set of the boolean variables $s_{x,a}$ for $x \in X$ and $a \in A$, i.e.,

$$s_{x,a} = \begin{cases} 1, & \text{if } a = s(x); \\ 0, & \text{if } a \neq s(x). \end{cases} \tag{2.112}$$

Lemma 2.47 *For a fixed strategy s the values σ_x^γ, $x \in X$ determine the unique optimal basic solution of the following linear programming problem:*
Maximize

$$\varphi_{x_0}^s(\sigma) = \sigma_{x_0} \tag{2.113}$$

subject to

$$\sigma_x - \gamma \sum_{y \in X} \sum_{a \in A(x)} s_{x,a}\, p_{x,y}^a\, \sigma_y \leq \sum_{a \in A(x)} s_{x,a}\, \mu_{x,a}, \quad \forall x \in X; \tag{2.114}$$

where

$$\mu_{x,a} = \sum_{a \in A(x)} c_{x,y}^a\, p_{x,y}^a, \quad \forall x \in X$$

and

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

Proof For a fixed strategy s the solution of the system of linear equations

$$\sigma_x - \gamma \sum_{y \in X} \sum_{a \in A(x)} s_{x,a}\, p_{x,y}^a\, \sigma_y = \sum_{a \in A(x)} s_{x,a}\, \mu_{x,a}, \quad \forall x \in X \tag{2.115}$$

uniquely determines σ_x, $\forall x \in X$. Thus, the problem of maximization of the objective function (2.113) subject to (2.115) for a fixed strategy s has a unique feasible solution which is an optimal one. This implies that if for fixed s we consider the problem: Maximize (2.113) subject to

$$\sigma_x - \gamma \sum_{y \in X} \sum_{a \in A(x)} s_{x,a}\, p_{x,y}^a\, \sigma_y \leq \sum_{a \in A(x)} s_{x,a}\, \mu_{x,a}, \quad \forall x \in X \tag{2.116}$$

then it has the same optimal solution as the problem (2.113), (2.115). Moreover, if in the problem (2.113), (2.116) we vary the boolean variables $s_{x,y}$ and take the maximum with respect to σ and the minimum with respect to s then we obtain the optimal strategy for the control problem. □

Using the lemma above we can prove the following theorem.

Theorem 2.48 *Let $\alpha^*_{x,a}$, β^*_y $(x \in X,\ a \in A)$ be a basic optimal solution of the following linear programming problem:*
Minimize

$$\phi_{x_0}(\alpha, \beta) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a}\, \alpha_{x,a} \tag{2.117}$$

subject to

$$\begin{cases} \beta_y - \gamma \displaystyle\sum_{x \in X} \sum_{a \in A(x)} p^a_{x,y}\, \alpha_{x,a} = 1, \quad y = x_0; \\ \beta_y - \gamma \displaystyle\sum_{x \in X} \sum_{a \in A(x)} p^a_{x,y}\, \alpha_{x,a} = 0, \quad \forall y \in X \backslash \{x_0\}; \\ \displaystyle\sum_{a \in A(x)} \alpha_{x,a} = \beta_x, \quad \forall x \in X; \\ \beta_y \geq 0, \quad \forall y \in X; \quad \alpha_{x,a} \geq 0, \quad \forall x \in X,\ a \in A(x). \end{cases} \tag{2.118}$$

Then the optimal stationary strategy s^* for the discounted Markov decision problem is determined as follows:

$$s^*_{x,a} = \begin{cases} 1, & \text{if } \alpha^*_{x,a} \neq 0; \\ 0, & \text{if } \alpha^*_{x,a} = 0. \end{cases} \tag{2.119}$$

Proof According to Lemma 2.47 the optimal stationary strategy s^* corresponds to the optimal solution of the problem (2.113), (2.114).

In a similar way as in the proof of Lemma 2.37 here we have that a stationary strategy s corresponds to an extreme point of the set of solutions of the following system

$$\begin{cases} \displaystyle\sum_{a \in A(x)} s_{x,a} = 1, \quad \forall x \in X; \\ s_{x,a} \geq 0, \quad \forall x \in X,\ a \in A. \end{cases} \tag{2.120}$$

Therefore, if we dualize (2.113), (2.120) with respect to σ_x for a fixed strategy s then we obtain the following optimization problem:
Minimize

$$\phi_{x_0}(s, \beta) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} s_{x,a}\, \beta_x \tag{2.121}$$

subject to

$$\begin{cases} \beta_y - \gamma \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, s_{x,a}\, \beta_x = 1, \quad y = x_0; \\ \beta_y - \gamma \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, s_{x,a}\, \beta_x = 0, \quad \forall y \in X \backslash \{x_0\}; \\ \beta_y \geq 0, \quad \forall y \in X. \end{cases} \tag{2.122}$$

Now if we minimize (2.113) with respect to $s_{x,a}$ and β_x and in (2.120) we take into account the following restriction

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

then we obtain the problem:
Minimize

$$\phi_{x_0}(s, \beta) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} s_{x,a}\, \beta_x \tag{2.123}$$

subject to

$$\begin{cases} \beta_y - \gamma \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, s_{x,a}\, \beta_x = 1, \quad y = x_0; \\ \beta_y - \gamma \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, s_{x,a}\, \beta_x = 0, \quad \forall y \in X \backslash \{x_0\}; \\ \sum\limits_{a \in A(x)} s_{x,a} = 1, \quad \forall x \in X; \\ \beta_y \geq 0, \quad \forall y \in X; \quad s_{x,a} \geq 0, \quad \forall x \in X,\ a \in A(x). \end{cases} \tag{2.124}$$

This bilinear programming problem can be easily reduced to a linear programming problem (2.117), (2.118) using the following elementary transformations: We change in (2.124) the restrictions $\sum_{a \in A(x)} s_{x,a} = 1,\ \forall x \in X$ by $\sum_{a \in A(x)} s_{x,a}\, \beta_x = \beta_x,\ \forall x \in X$ and then we introduce the notations

$$\alpha_{x,a} = s_{x,a}\, \beta_x, \quad \forall x \in X,\ a \in A(x). \tag{2.125}$$

If ${\alpha^a_{x,y}}^*$, β^*_y $(x, y \in X, a \in A)$ is a basic optimal solution of the linear programming problem (2.117), (2.118) then by using (2.125) we obtain $s^*_{x,a}$ according to (2.119). □

Based on Theorem 2.48 we can propose the following algorithm for determining the solution of the Markov decision problem.

Algorithm 2.49 Determining the Optimal Stationary Strategy for the Discounted Markov Decision Problem

(1) Formulate the linear programming problem (2.117), (2.118);
(2) Determine a basic optimal solution $\alpha^*_{x,a}$ $(x \in X,\ a \in A)$, β^*_y $(y \in X)$ of the problem (2.117), (2.118) and determine $s^*_{x,x}$ according to (2.119).

The results described above allow us to determine the stationary strategy for the discounted Markov decision problem in the case if the starting state x_0 is fixed. In the general case, if it is necessary to find the optimal stationary strategy for an arbitrary starting state $x \in X$ then we can use the following results.

Lemma 2.50 *For a fixed strategy s the values σ^{γ}_x, $x \in X$ determine the unique optimal basic solution of the following linear programming problem:*
Maximize

$$\varphi^s(\sigma) = \sum_{x \in X} \sigma_x$$

subject to

$$\sigma_x - \gamma \sum_{y \in X} \sum_{a \in A(x)} s_{x,a}\, p^a_{x,y}\, \sigma_y \ \le \sum_{a \in A(x)} s_{x,a}\, \mu_{x,a}, \quad \forall x \in X,$$

where

$$\mu_{x,a} = \sum_{a \in A(x)} c^a_{x,y}\, p^a_{x,y}, \quad \forall x \in X$$

and

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

The proof of this lemma is similar to the proof of Lemma 2.47.
Using this lemma we can prove the following theorem:

Theorem 2.51 *Let $\alpha^*_{x,a}$, β^*_y $(x \in X,\ y \in X,\ a \in A)$ be a basic optimal solution of the following linear programming problem:*
Minimize

$$\phi(\alpha, \beta) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha_{x,a} \tag{2.126}$$

subject to

$$\begin{cases} \beta_y - \gamma \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, \alpha_{x,a} = 1, \quad \forall y \in X; \\ \sum\limits_{a \in A(x)} \alpha_{x,a} = \beta_x, \ \forall x \in X; \\ \beta_y \geq 0, \quad \forall y \in X; \quad \alpha_{x,a} \geq 0, \quad \forall x \in X, \ a \in A(x). \end{cases} \tag{2.127}$$

Then the optimal stationary strategy s^* for the discounted Markov decision problem is determined according to (2.119).

The proof of this theorem is identical to the prove of Theorem 2.48; here we have to apply Lemma 2.50 instead of Lemma 2.47.

It is easy to observe that the constraints $\beta_y \geq 0, \ \forall y \in X$, in (2.127) are redundant. Therefore, we can eliminate $\beta_x, \forall x \in X$, from (2.127) introducing the expressions $\sum_{a \in A(x)} \alpha_{x,a} = \beta_x$ for $x \in X$ in the first group of the constraints. After that from Theorem 2.51 we obtain the following corollary.

Corollary 2.52 *Let $\alpha^*_{x,a}$ $(x \in X, \ y \in X, \ a \in A)$ be a basic optimal solution of the following linear programming problem:*
Minimize

$$\phi(\alpha) = \sum_{x \in X} \sum_{a \in A(x)} \mu_{x,a} \alpha_{x,a} \tag{2.128}$$

subject to

$$\begin{cases} \sum\limits_{a \in A(x)} \alpha_{x,a} - \gamma \sum\limits_{x \in X} \sum\limits_{a \in A(x)} p^a_{x,y}\, \alpha_{x,a} = 1, \quad \forall y \in X; \\ \alpha_{x,a} \geq 0, \quad \forall x \in X, \ a \in A(x). \end{cases} \tag{2.129}$$

Then the optimal stationary strategy s^* for the discounted Markov decision problem is determined as follows:

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

2.6.1 A Dual Linear Programming Model for the Discounted Markov Decision Problem

We formulate the dual linear programming model for the discounted Markov decision problem using the problem (2.128), (2.129). Applying the duality linear programming theorems to this problem we obtain the following result:

Theorem 2.53 *Let σ_x^* $(x \in X)$ be the optimal solution of the linear programming problem:*
Maximize

$$\varphi(\sigma) = \sum_{x \in X} \sigma_x \tag{2.130}$$

subject to

$$\sigma_x - \gamma \sum_{y \in X} p_{x,y}^a \sigma_y \leq \mu_{x,a}, \quad \forall x \in X,\ a \in A(x). \tag{2.131}$$

Then σ_x^* for $x \in X$ represents the optimal discounted expected total costs for the problem on the network with starting states $x \in X$. An optimal stationary strategy can be found by fixing $s^* : X \to A$ such that $s^*(x) = a \in A^*(x)$, $\forall x \in X$, where $A^*(x) = \{a \in A(x) \,|\, \sigma_x - \gamma \sum_{y \in X} p_{x,y}^a \sigma_y = 0\}$.

Thus, the solution of the discounted Markov decision problem can be found by solving the dual linear programming problem (2.130), (2.131).

2.7 An Iterative Algorithm for Discounted Markov Decision Processes and Stochastic Control Problems

To determine the optimal discounted costs and the corresponding optimal strategy in the Markov processes with discounted costs we shall use the following system of equations with respect to $\sigma_{x_1}, \sigma_{x_2}, \ldots, \sigma_{x_n}$:

$$\sigma_{x_i} = \min_{a \in A(x_i)} \left[\mu_{x_i,a} + \gamma \sum_{x_j \in X} p_{x_i,x_j}^a \sigma_{x_j} \right], \quad i = 1, 2, \ldots, n.$$

According to Theorem 2.53 this system of equations has a solution. Below we describe an iterative algorithm for determining the solution of this system of equations and finding the optimal stationary strategies of the discounted Markov decision problem.

Algorithm 2.54 Determining the Optimal Stationary Strategies for the Discounted Markov Decision Problem

Preliminary step (Step 0): Fix an arbitrary stationary strategy

$$s^0 : x_i \to a \in A(x_i) \ \text{ for } \ x_i \in X.$$

General step (Step k, $k > 0$): Calculate

$$\mu_{x_i,s^{k-1}(x_i)} = \sum_{y \in X(x_i)} p_{x_i,y}^{s^{k-1}(x_i)} c_{x_i,y}^{s^{k-1}(x_i)}$$

for every $x_i \in X$. Then solve the system of linear equations

$$\sigma_{x_i} = \mu_{x_i,s^{k-1}(x_i)} + \gamma \sum_{x_j \in X} p_{x_i,x_j}^{s^{k-1}(x_i)} \sigma_{x_j}, \quad i = 1, 2, \ldots, n$$

and find the solution $\sigma_{x_1}^{k-1}, \sigma_{x_2}^{k-1}, \ldots, \sigma_{x_n}^{k-1}$. After that determine a new strategy

$$s^k : x_i \to a \in A(x_i) \text{ for } x_i \in X,$$

where

$$s^k(x_i) = \operatorname*{argmin}_{a \in A(x_i)} \left[\mu_{x_i,a} + \gamma \sum_{x_j \in X} p_{x_i,x_j}^a \sigma_{x_i}^{k-1} \right], \quad i = 1, 2, \ldots, n.$$

Check if the following condition holds

$$s^k(x_i) = s^{k-1}(x_i), \quad \forall x_i \in X. \tag{2.132}$$

If the condition (2.132) holds then fix

$$s^* = s^k; \quad \sigma_{x_i}^* = \sigma_{x_i}^k, \quad \forall x_i \in X$$

as the optimal solution of the problem; otherwise go to the next step $k + 1$.

This algorithm can be specified for a stochastic control problem with a discounted cost criterion. The correctness and the convergence of this iterative algorithm can be derived from the results described above and the results from [32, 112, 128, 136].

Algorithm 2.55 Determining the Optimal Stationary Strategies for the Discounted Stochastic Control Problem

We consider the discounted control problem on the network (G, X_1, X_2, c, p) with a given discount factor γ. The dynamics of the system is described by a directed graph $G = (X, E)$ with the set of controllable states X_C and the set of uncontrollable states X_N. In addition we assume that the probability function $p : E_2 \to [0, 1]$ and the cost function $c : E \to \mathbb{R}$ are given.

Preliminary step (Step 0): Fix an arbitrary stationary strategy

$$s^0 : x_i \to x_j \in X(x_i) \text{ for } x_i \in X_C.$$

General step (Step k, $k > 0$): Determine the probability matrix $P^{s^{k-1}} = (p^{s^{k-1}}_{x_i,x_j})$, where

$$p^{s^{k-1}}_{x_i,x_j} = \begin{cases} p_{x_i,x_j}, & \text{if } x_i \in X_N \text{ and } (x_i,x_i) \in E_N; \\ 1, & \text{if } x_i \in X_C \text{ and } x_j = s^{k-1}(x_i); \\ 0, & \text{if } x_i \in X_C \text{ and } x_j \neq s^{k-1}(x_i). \end{cases}$$

Then calculate

$$\mu_{x_i,s^{k-1}(x_i)} = \sum_{y \in X(x_i)} p^{s^{k-1}(x_i)}_{x_i,y} c^{s^{k-1}(x_i)}_{x_i,y}$$

for every $x_i \in X$ and solve the system of linear equations

$$\sigma_{x_i} = \mu_{x_i,s^{k-1}(x_i)} + \gamma \sum_{x_j \in X} p^{s^{k-1}(x_i)}_{x_i,x_j} \sigma_{x_j}, \quad i = 1, 2, \ldots, n$$

and find the solution $\sigma^{k-1}_{x_1}, \sigma^{k-1}_{x_2}, \ldots, \sigma^{k-1}_{x_n}$. After that determine a new strategy

$$s^k : x_i \to a \in A(x_i) \text{ for } x_i \in X_C,$$

where

$$s^k(x_i) = \operatorname*{argmin}_{a \in A(x_i)} \left[\mu_{x_i,a} + \gamma \sum_{x_j \in X} p^a_{x_i,x_j} \sigma^{k-1}_{x_i} \right], \quad \forall x_i \in X_C.$$

Check if the following condition holds

$$s^k(x_i) = s^{k-1}(x_i), \quad \forall x_i \in X_C. \tag{2.133}$$

If the condition (2.133) holds then fix

$$s^* = s^k; \quad \sigma^*_{x_i} = \sigma^k_{x_i}, \ \forall x_i \in X$$

as the optimal solution of the problem; otherwise go to the next step $k + 1$.

This algorithm finds the optimal stationary strategy for an arbitrary stochastic control problem. In the case when $X = X_C$ ($X_N = \emptyset$) we obtain an iterative algorithm for deterministic discounted control problems.

2.8 Determining the Optimal Expected Total Cost for Markov Decision Problems with a Stopping State

The algorithms proposed in the previous sections determine the optimal stationary strategies for discounted Markov decision problems in the case if the discount factor γ satisfies the condition $0 < \gamma < 1$. If $\gamma = 1$ then the expected total cost in these problems may not exist. Here we study a class of unichain decision problems for which γ may be equal to 1 and the expected total cost exists. Moreover, we can see that for some problems γ may be an arbitrary positive value. For the considered problems we show how to determine the optimal expected total cost using a linear programming approach and iterative procedures. To ensure the existence of the expected total cost in these problems we assume that for the dynamical system there exists a state in which transitions stop as soon as this state is reached [89]. Furthermore, we describe algorithms for determining optimal strategies in such problems.

2.8.1 Problem Formulation and a Linear Programming Approach

Let (X, A, p, c) be a Markov decision process with a finite set of states X, a finite set of actions A, the probability function $p : A \times X \times X \to \mathbb{R}^+$ that satisfies the condition $\sum_{y \in X} p^a_{x,y} = 1, \ \forall a \in A$ and the transition cost function $c : A \times X \times X \to \mathbb{R}$. In addition a discount factor γ for the Markov decision process is given, where $0 < \gamma \leq 1$. We consider the problem of determining the stationary strategy with minimal expected total cost for unichain Markov processes in the case if the dynamical system stops transitions in a given state $z \in X$. At first we assume that the Markov process is perfect. Moreover, we assume that for an arbitrary fixed action in this decision process the state $z \in X$ is an absorbing state. Obviously, in this case for $0 < \gamma < 1$ the optimal expected total costs σ_x and the optimal stationary strategy for an arbitrary starting state $x \in X \backslash \{z\}$ can be found using the linear programming models (2.128), (2.129) and (2.130), (2.131) considering $c^a_{z,z} = 0, \ \forall a \in A(z)$. If the optimal strategy s^* is found then we have only to fix $s^*(x)$ for $x \in X \backslash \{z\}$ because z is the stopping state. In this case the expected total cost for a given starting state σ_{x_0} can be found by solving the linear programming problem (2.117), (2.118). Now we can see that the considered linear programming models can be used for determining the solution of the decision problem with an absorbing stopping state $z \in X$ in the case $\gamma = 1$ if $c^a_{z,z} = 0, \ \forall a \in A(x)$. Indeed, for a fixed strategy s the rank of the matrix $(I - P^s)$ for the unichain process is equal to $|X| - 1$ and the system of equations $(I - P^s)\sigma = \mu^s$ has a unique solution if we put $\sigma_z = 0$. Thus, for unichain processes with absorbing state $z \in X$ the system of equations

$$\begin{cases} (I - P^s)\sigma = \mu^s; \\ \sigma_z = 0 \end{cases}$$

has a unique solution if $c^a_{z,z} = 0, \ \forall a \in A(z)$.

The properties mentioned above allow us to conclude that for a unichain decision problem with $0 < \gamma \leq 1$ the following lemma holds.

Lemma 2.56 *A stationary strategy s^* is optimal if and only if it corresponds to an optimal solution σ^*, s^* of the following mixed integer bilinear programming problem:*
Maximize

$$\varphi_{x_0}(\sigma, s) = \sigma_{x_0} \tag{2.134}$$

subject to

$$\begin{cases} \sigma_x - \gamma \sum\limits_{y \in X} \sum\limits_{a \in A(x)} s_{x,a}\, p^a_{x,y}\, \sigma_y \leq \sum\limits_{a \in A(x)} s_{x,a}\, \mu_{x,a}, \quad \forall x \in X \backslash \{z\}; \\ \sigma_z = 0; \\ \sum\limits_{a \in A(x)} s_{x,a} = 1, \quad \forall x \in X \backslash \{z\}; \\ s_{x,a} \in \{0, 1\}, \quad \forall x \in X \backslash \{z\},\ a \in A(x), \end{cases} \tag{2.135}$$

where

$$\mu_{x,a} = \sum_{y \in X} p^a_{x,y}\, c^a_{x,y}.$$

Note that in (2.134), (2.135) the boolean variables $s_{x,a}$ for $x \in X \backslash \{z\}$, $a \in A(x)$ correspond to a strategy $s : X \backslash \{z\} \to X$, where $s_{x,a} = 1$ if $s(x) = a$ and $s_{x,a} = 0$ if $s(x) \neq a$. Based on this lemma, we can prove the following theorem.

Theorem 2.57 *Let ${\alpha_{x,a}}^*$, β^*_y ($x \in X \backslash \{z\}$, $y \in X \backslash \{z\}$, $a \in A$) be a basic optimal solution of the following linear programming problem:*
Minimize

$$\phi_{x_0}(\alpha, \beta) = \sum_{x \in X \backslash \{z\}} \sum_{a \in A(x)} \mu_{x,a}\, \alpha_{x,a} \tag{2.136}$$

subject to

$$\begin{cases} \beta_y - \gamma \sum\limits_{x \in X \backslash \{z\}} \sum\limits_{a \in A(x)} p^a_{x,y}\, \alpha_{x,a} \geq 1, \quad y = x_0; \\ \beta_y - \gamma \sum\limits_{x \in X \backslash \{z\}} \sum\limits_{a \in A(x)} p^a_{x,y}\, \alpha_{x,a} \geq 0, \quad \forall y \in X \backslash \{x_0, z\}; \\ \sum\limits_{a \in A(x)} \alpha_{x,a} = \beta_x, \quad \forall x \in X \backslash \{z\}; \\ \beta_y \geq 0, \quad \forall y \in X \backslash \{z\}; \quad \alpha_{x,a} \geq 0, \quad \forall x \in X \backslash \{z\},\ a \in A(x). \end{cases} \tag{2.137}$$

Then the optimal stationary strategy s^* for the discounted unichain decision problem with absorbing state $z \in X$ is determined as follows:

$$s^*_{x,a} = \begin{cases} 1, & \text{if } \alpha^*_{x,a} \neq 0; \\ 0, & \text{if } \alpha^*_{x,a} = 0. \end{cases} \tag{2.138}$$

The proof of Theorem 2.57 is obtained in the same way as Theorem 2.48. Lemma 2.56 and Theorem 2.57 differ from Lemma 2.47 and Theorem 2.48, respectively, only in a single restriction in the systems (2.135) and (2.137). These systems are obtained from (2.114) and (2.114), respectively, by deleting the constraints that correspond to the absorbing state z. In the proof of Theorem 2.57 we have only to assume that the expected total cost for the problem with an absorbing stopping state exists.

Remark 2.58 The values σ_x, $\forall x \in X$ for a unichain Markov decision problem with stopping state z with $c^a_{z,z} = 0$, $\forall a \in A(z)$ and $\gamma = 1$ coincide with the values ε_x, $\forall x \in X$ for an zero average cost Markov decision problem.

Based on Theorem 2.57 the optimal stationary strategy of the problem with stopping state can be found by using the following algorithm.

Algorithm 2.59 Determining the Optimal Stationary Strategy for a Markov Decision Problem with Stopping State

(1) Formulate the linear programming problem (2.136), (2.137);
(2) Determine a basic optimal solution $\alpha^*_{x,a}$ $(x \in X\backslash\{z\},\ a \in A)$, β^*_y $(y \in X\backslash\{z\})$ of the problem (2.136), (2.137) and determine $s^*_{x,x}$ according to (2.138).

Remark 2.60 Theorem 2.57 and Algorithm 2.59 are also valid for an arbitrary Markov decision problem with a stopping state z in the case $\gamma \geq 1$ if the cost function $c : X \times X \times A \to \mathbb{R}$ is strict positive and there exists a strategy s that induces a unichain process with a stopping absorbing state z. Thus, Theorem 2.57 in the case $\gamma \geq 1$ gives necessary and sufficient conditions for determining the optimal stationary strategies in the discounted decision problem with positive costs and given stopping state z.

If in the considered decision problem it is necessary to determine the optimal stationary strategies for an arbitrary starting state $x \in X\backslash\{z\}$ then we can use the following result.

Theorem 2.61 *Let $\alpha^*_{x,a}$ $(x \in X\backslash\{z\},\ a \in A)$ be a basic optimal solution of the following linear programming problem:*
Minimize

$$\phi(\alpha) = \sum_{x \in X\backslash\{z\}} \sum_{a \in A(x)} \mu_{x,a}\alpha_{x,a} \tag{2.139}$$

subject to

$$\begin{cases} \sum\limits_{a\in A(y)} \alpha_{y,a} - \gamma \sum\limits_{x\in X\setminus\{z\}} \sum\limits_{a\in A(x)} p^a_{x,y}\,\alpha_{x,a} \geq 1, \quad \forall y \in X\setminus\{z\}; \\ \alpha_{x,a} \geq 0, \quad \forall x \in X\setminus\{z\},\ a \in A(x). \end{cases} \tag{2.140}$$

Then the optimal stationary strategy s^* for the discounted Markov decision problem with an arbitrary starting state $x \in X\setminus\{z\}$ and given stopping state z is determined as follows:

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

The proof of this theorem can be obtained by using the following lemma.

Lemma 2.62 *A stationary strategy s^* is optimal if and only if it corresponds to an optimal solution σ^*, s^* of the following mixed integer bilinear programming problem:*
Maximize

$$\varphi(\sigma, s) = \sum_{x\in X} \sigma_x$$

subject to (2.135).

If for the linear programming problem (2.139), (2.140) we construct the dual model in the same way as for the previous problems then we obtain the following result.

Theorem 2.63 *Let σ^*_x $(x \in X)$ be the optimal solution of the linear programming problem:*
Maximize

$$\varphi(\sigma) = \sum_{x\in X} \sigma_x \tag{2.141}$$

subject to

$$\sigma_x - \gamma \sum_{y\in X} p^a_{x,y}\sigma_y \leq \mu_{x,a}, \quad \forall x \in X\setminus\{z\},\ a \in A(x), \tag{2.142}$$

where $0 < \gamma \leq 1$. Then σ^*_x for $x \in X$ represents the optimal discounted expected total cost for the problem with starting states $x \in X$. An optimal stationary strategy can be found by fixing $s^* : X\setminus\{z\} \to A$ such that $s^*(x) = a \in A^*(x)$, $\forall x \in X\setminus\{z\}$, where $A^*(x) = \{a \in A(x) \mid \sigma_x - \gamma \sum_{y\in X} p^a_{x,y}\sigma_y = \mu_{x,a}\}$.

We can obtain an iterative algorithm for the problems with a stopping state from the algorithm for a discounted Markov decision problem from Sect. 2.7 if at each iteration of the algorithm we solve the system of linear equations

$$\begin{cases} \sigma_z = 0; \\ \sigma_{x_i} = \mu_{x_i, s^{k-1}(x_i)} + \gamma \sum\limits_{x_j \in X \setminus \{z\}} p_{x_i, x_j}^{s^{k-1}(x_i)} \sigma_{x_j}, \quad \forall x_i \in X \setminus \{z\} \end{cases}$$

instead of the system of linear equations

$$\sigma_{x_i} = \mu_{x_i, s^{k-1}(x_i)} + \gamma \sum_{x_j \in X} p_{x_i, x_j}^{s^{k-1}(x_i)} \sigma_{x_j}, \quad \forall x_i \in X.$$

Thus, if in the general step of the iterative algorithm from the previous section we replace this system of the equations by the system of equations written above we obtain the iterative algorithm for the problem with an absorbing state. Now let us show how to solve the unichain Markov decision problem with a given stopping state $z \in x$ if z is not an absorbing state but is a positive recurrent state of the Markov process induced by an arbitrary stationary strategy. In this case the problem can be reduced to the case with an absorbing stopping state if we make the following minor transformations in the unichain Markov decision process: For an arbitrary action $a \in A(z)$ we set $p_{z,y}^a = 0$, $\forall y \in X \setminus \{z\}$; $p_{z,z}^a = 1$. Obviously, after such a transformation of the unichain decision process we obtain the optimal stationary strategies of the problem if the state z is reached.

2.8.2 *Optimality Conditions for the Control Problem on Network with a Stopping State*

Consider a discounted control problem for the decision network (G, X_C, X_N, c, p) with a given stopping state $z \in X$ and a given discount factor γ, $0 < \gamma \leq 1$. Then on the basis of Theorem 2.42 the following result can be proved.

Theorem 2.64 *Assume that in G an arbitrary stationary strategy $s : x \to X(x)$ for $x \in X_C$ generates a subgraph $G_s = (X, E_s \cup E_N)$ where the vertex z can be reached from arbitrary $x \in X \setminus \{z\}$. Then the linear programming problem:*
Minimize

$$\phi(\alpha, \beta) = \sum_{x \in X_C} \sum_{y \in X(x)} c_{x,y} \alpha_{x,y} + \sum_{x \in X_N} \mu_x \beta_x \tag{2.143}$$

subject to

$$\begin{cases} \sum\limits_{x\in X(y)} \alpha_{y,x} - \gamma \sum\limits_{x\in X_C^-(y)} \alpha_{x,y} - \gamma \sum\limits_{x\in X_N^-(y)} p_{x,y}\,\beta_x = 1, \quad y \in X_C\backslash\{z\}; \\ \beta_y - \gamma \sum\limits_{x\in X_C^-(y)} \alpha_{x,y} - \gamma \sum\limits_{x\in X_N^-(y)} p_{x,y}\,\beta_x = 1, \quad y \in X_N\backslash\{z\}; \\ \beta_x \geq 0,\ \forall x \in X_N; \quad \alpha_{x,y} \geq 0, \quad \forall x \in X_C,\ y \in X(x), \end{cases} \tag{2.144}$$

has a solution. If α^*, β^* *is an arbitrary basic solution of the problem* (2.143), (2.144) *then the optimal stationary strategy* s^* *for the discounted control problem with a stopping state* z *can be found by fixing* $s^*_{x,y} = 1$ *for* $x \in X_C,\ y \in X(x)$ *if* $\alpha^*_{x,y} > 0$, *and* $s_{x,y} = 0^*$ *in the other case.*

It is easy to observe that the problem (2.143), (2.144) is obtained from the problem (2.107), (2.108) by deleting the restriction that corresponds to a stopping state z. Thus, if the conditions of the theorem hold then the problem has a solution for arbitrary $\gamma \in (0, 1]$.

The optimality conditions for control problems on networks with a stopping state can be derived if we consider the dual model for the problem (2.143), (2.144) or from Theorem 2.63. If we specify this theorem for the problem on networks then we obtain the following result.

Theorem 2.65 *Let* (G, X_C, X_N, c, p) *be a perfect decision network with a given stopping state* z *and a given discount factor* $\gamma\ (0 < \gamma \leq 1)$, *where the function* $c : E \to \mathbb{R}$ *is strictly positive. Then the optimal expected discounted total cost* σ^*_x *of the control problem on the decision network exists for an arbitrary fixed starting state* $x \in X$. *The values* σ^*_x, *for* $x \in X\backslash\{z\}$ *can be found by solving the following linear programming problem:*
Maximize

$$\varphi(\sigma) = \sum_{x\in X} \sigma_x \tag{2.145}$$

subject to

$$\begin{cases} \sigma_x - \gamma\sigma_y \leq c_{x,y}, \ \forall x \in X_C\backslash\{x\}, \ y \in X(x); \\ \sigma_x - \gamma \sum\limits_{y\in X} p_{x,y}\sigma_y \leq \mu_x, \quad \forall x \in X_N\backslash\{z\} \end{cases} \tag{2.146}$$

and the optimal stationary strategy can be determined by fixing $s^* : X\backslash\{z\} \to A$ *such that* $s^*(x) = y \in X^*(x),\ \forall x \in X_C\backslash\{z\}$, *where* $X^*(x) = \{y \in X(x) \mid \sigma_x - \gamma\sigma_y = c_{x,y}\}$.

Corollary 2.66 *Let* (G, X_C, X_N, c, p) *be a perfect decision network that satisfies the conditions of* Theorem 2.65. *Then for an arbitrary* $\gamma \in (0, 1]$ *there exist the values* σ_x^* *for* $x \in X$ *that satisfy the conditions:*

(1) $c_{x,y} + \gamma\sigma_y^* - \sigma_x^* \geq 0, \ \forall x \in X_C \backslash \{x\}, \ y \in X(x)$;

(2) $\min\limits_{y \in X(x)} (c_{x,y} + \gamma\sigma_y^* - \sigma_x^*) = 0, \ \forall x \in X_C$;

(3) $\mu_x + \gamma \sum\limits_{y \in X} p_{x,y}\sigma_y^* - \sigma_x^* = 0, \quad \forall x \in X_N \backslash \{z\}$;

An optimal stationary strategy of the optimal control problem on the network (G, X_C, X_N, c, p) *with stopping state* z *can be found by fixing* $s^* : X\backslash\{z\} \to A$ *such that* $s^*(x) = y \in X^*(x), \ \forall x \in X_C\backslash\{z\}$, *where* $X^*(x) = \{y \in X(x) \mid c_{x,y} + \gamma\sigma_y^* - \sigma_x^* = 0\}$.

Remark 2.67 The control problem on the network with a given stopping state z in the case $X_N = 0$, $\gamma = 1$ becomes the problem of determining in G the minimum cost paths from $x \in X$ to z. If G is an acyclic graph with sink vertex then the problem has a solution for an arbitrary $\gamma > 0$.

2.8.3 A Dynamic Programming Algorithm for Solving Deterministic Non-stationary Control Problems on Networks

As we have noted the deterministic stationary control problem on networks with fixed stopping state z corresponds to the case $X_N = \emptyset$ and the solution can be found by using linear programming models (2.143), (2.144) and (2.145), (2.146). In this section we show that this problem can be solved for the non-stationary case using a dynamic programming method. We describe an algorithm for finding the solution of the deterministic control problem on the network when the costs on the edges may depend on time. So, we assume that $X_N = \emptyset$ and in the network to each directed edge $e = (x, y) \in E$ the cost function $c_e(t)$ that depends on t is associated. This means that if the system makes a transition from the state $x = x(t)$ to the state $y = x(t+1)$ then the cost is $c_{x,y}(t)$. Thus, the problem in this case is formulated in the following way:

For a given time-moment $\bar{t}$ and fixed starting and stopping states $x_0, x_f \in X$ it is necessary to determine in G a sequence of the system's transitions $(x(0), x(1))$, $(x(1), x(2)), \ldots, (x(\bar{t}-1), x(\bar{t}))$, which transfers the system $\mathbb{L}$ from a starting state $x_0 = x(0)$ to a stopping state $x_f = x(\bar{t})$ such that the total cost

$$F_{x_0 x_f}(\bar{t}) = \sum_{t=0}^{\bar{t}-1} c_{(x(t),x(t+1))}(t)$$

of the system's transitions by a trajectory

$$x_0 = x(0), x(1), x(2), \ldots, x(\bar{t}) = x_f$$

is minimal, where $(x(t), x(t+1)) \in E, \quad t = 0, 1, 2, \ldots, \bar{t} - 1$.

We describe the dynamic programming algorithm for solving this problem. Denote by

$$F^*_{x_0,x_f}(\bar{t}) = \min_{x_0=x(0),x(1),\ldots,x(\bar{t})=x_f} \sum_{t=0}^{\bar{t}-1} c_{(x(t),x(t+1))}(t)$$

the minimal total cost of the system's transition from x_0 to x_f with $\bar{t}$ stages, where $F^*_{x_0,x_f}(0) = 0$ in the case $x_0 = x_f$ and $F^*_{x_0x_f}(\bar{t}) = \infty$ if x_f cannot be reached from x_0 by using $\bar{t}$ transitions.

If we introduce the values $F^*_{x_0x(t)}(t)$ for $t = 0, 1, 2, \ldots, \bar{t} - 1$ then it is easy to observe that for $F^*_{x_0x(t)}(t)$ the following recursive formula can be gained:

$$F^*_{x_0x(t)}(t) = \min_{x(t-1) \in X_G^-(x(t))} \left\{ F^*_{x_0,x(t-1)}(t-1) + c_{(x(t-1),x(t))}(t-1) \right\},$$

where

$$F^*_{x_0,x(0)}(0) = 0$$

and

$$X_G^-(y) = \{x \in X \mid e = (x, y) \in E\}.$$

Based on this recursive formula we can tabulate the values $F^*_{x_0x(t)}(t), t = 1, 2, \ldots, \bar{t}$ for every $x(t) \in X$. These values and the solution of the problem can be found using $O(|X|^2\bar{t})$ elementary operations (here we do not take into account the number of operations for calculating the values of the functions $c_e(t)$ for a given t).

The tabulation process should be organized in such a way that for every vertex $x = x(t)$ at a given moment in time t it is determined not only the cost $F^*_{x_0x(t)}(t)$ but also the state $x^*(t-1)$ at the previous time-moments for which

$$\begin{aligned} F^*_{x_0,x(t)}(t) &= F^*_{x_0x^*(t-1)} + c_{(x^*(t-1),x(t))}(t-1) \\ &= \min_{x(t-1) \in X_G^-(x(t))} \{F^*_{x_0,x(t-1)} + c_{(x(t-1),x(t))}(t-1)\}. \end{aligned}$$

So, if to each x at the time-moments $t = 0, 1, 2, \ldots, \bar{t}$ we associate the labels $(t, x(t), F^*_{x_0x(t)}, x^*(t-1))$, then the corresponding table allows us to find the optimal trajectory successively starting from the final position, $x_f = x^*(\bar{t}), x^*(\bar{t} - 1), \ldots, x^*(1), x^*(0) = x_0$. In the example given below all possible labels for every x and every t are represented in Table 2.1.

Table 2.1 The values $F^*_{x_0,x(t)}$ and $x^*(t-1)$

t	x, F^*	0	1	2	3	4	5
0	$F^*_{x_0,x(0)}$	0	∞	∞	∞	∞	∞
	$x^*(0-1)$	–	–	–	–	–	–
1	$F^*_{x_0,x(1)}$	∞	1	∞	1	∞	∞
	$x^*(0)$	–	0*	–	0	–	–
2	$F^*_{x_0,x(2)}$	∞	3	2	∞	5	2
	$x^*(1)$	–	3	1*	–	3	1
3	$F^*_{x_0,x(3)}$	∞	∞	5	6	4	3
	$x^*(2)$	–	–	1	2*	2	2
4	$F^*_{x_0,x(4)}$	∞	12	∞	11	8	6
	$x^*(3)$	–	3*	–	2	2	2
5	$F^*_{x_0,x(5)}$	∞	19	16	∞	21	16
	$x^*(4)$	–	3	1	–	3	1*

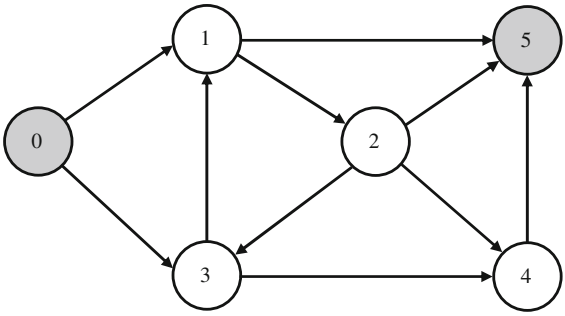

Fig. 2.13 The structure of the dynamic network

This problem can be extended to the case if the final state x_f should be reached at the moment of time $t(x_f)$ from a given interval $[\bar{t}_1, \bar{t}_2]$. If $\bar{t}_1 \neq \bar{t}_2$ then the problem can be reduced to $\bar{t}_2 - \bar{t}_1 + 1$ problems with $\bar{t} = \bar{t}_1$, $\bar{t} = \bar{t}_1 + 1$, $\bar{t} = \bar{t}_1 + 2, \ldots, \bar{t} = \bar{t}_2$, respectively; by comparing the minimal total costs of these problems we find the best one and $t(x_f)$.

An important case of the considered problem is if $\bar{t}_1 = 0$ and $\bar{t}_2 = \infty$. The solution of the problem with such a condition if the network may contain directed cycles has sense only for positive and non-decreasing cost functions $c_e(t)$ on the edges $e \in E$. Obviously, for this case we obtain $0 \leq t(x_f) \leq |X|$ and the problem can be solved in time $O(|X|^3)$ (the case with a free number of stages).

Example Let the dynamic network determined by the graph $G = (X, E)$ represented in Fig. 2.13 be given. The cost functions are the following:

$$c_{(0,1)}(t) = c_{(0,3)}(t) = c_{(2,5)}(t) = 1;$$
$$c_{(2,3)}(t) = c_{(3,1)}(t) = 2t; \quad c_{(3,4)}(t) = 2t + 2;$$
$$c_{(1,2)}(t) = c_{(2,4)}(t) = c_{(1,5)}(t) = t; \quad c_{(4,5)}(t) = 2t + 1.$$

We consider the problem of finding a trajectory in G from $x(0) = x_0 = 0$ to $x_f = 5$, where $T = 5$.

Using the recursive formula described above we get Table 2.1 with values $F^*_{x_0 x(t)}(t)$ and $x^*(t-1)$.

Starting from the final state $x_f = 5$ we find the optimal trajectory

$$5^* \leftarrow 1^* \leftarrow 3^* \leftarrow 2^* \leftarrow 1^* \leftarrow 0^*$$

with total cost $F_{x_0,x(5)}(5) = 16$.

The considered non-stationary control problem has been extended and generalized in [71, 79] as non linear minimum cost flow problems on dynamic networks. Algorithms based on *time-expanded network methods* for such a class of problems are described in [70, 71, 79, 93, 94].

2.9 Discrete Decision Problems with Varying Time of State's Transitions and Special Solution Algorithms

So far, in the control problems with average and discounted optimization cost criteria we have considered that the time between transitions in the control process is constant and it is equal to 1. We extend these problems and generalize these problems by assuming that the time of system's transition from one state to another in the decision process vary and it may be different from 1. Such a problem statement may be useful for studying and solving the decision models for the case of Semi-Markov processes. In this section we show that the deterministic problem with varying time of states' transitions can be reduced to the problem with a fixed unit time of system transitions from one state to another.

2.9.1 Problem Formulation

At first we formulate the control problem with an average cost optimization criterion when the transition time between the states is not constant.

Let the dynamical system $\mathbb{L}$ with a finite set of states $X \subseteq \mathbb{R}^n$ be given, where at every discrete moment of time $t = 0, 1, 2, \ldots$ the state of $\mathbb{L}$ is $x(t) \in X$. Assume, that the control of the system $\mathbb{L}$ at each time-moment $t = 0, 1, 2, \ldots$ for an arbitrary state $x(t)$ is realized by using the vector of control parameters $u(t) \in \mathbb{R}^m$ for which a feasible set $U_t(x(t))$ is given, i.e., $u(t) \in U_t(x(t))$. For arbitrary t and $x(t)$ on

$U_t\,(x(t))$ it is defined an integer function

$$\tau_{x(t)}:\ U_t\,(x(t)) \to \mathbb{N}$$

which represents to each control $u(t) \in U_t\,(x(t))$ an integer value $\tau_{x(t)}\,(u(t))$. This value expresses the time of system's transition from the state $x(t)$ to the state $x\big(t + \tau_{x(t)}\,(u(t))\big)$ if the control $u(t) \in U_t\,(x(t))$ has been applied at the moment t for a given state $x(t)$.

The dynamics of the system $\mathbb{L}$ is described by the following system of difference equations

$$\begin{cases} t_{j+1} &= t_j + \tau_{x(t_j)}\left(u(t_j)\right); \\ x(t_{j+1}) &= g_{t_j}\left(x(t_j), u(t_j)\right); \\ & u(t_j) \in U_{t_j}\left(x(t_j)\right); \\ & j = 0, 1, 2, \ldots, \end{cases}$$

where

$$x(t_0) = 0, \quad t_0 = 0$$

is a given starting state of the dynamical system $\mathbb{L}$. Here we suppose that the functions g_t and $\tau_{x(t)}$ are known and t_{j+1} and $x(t_{j+1})$ are determined uniquely by $x(t_j)$ and $u(t_j)$ at each step j.

Let $u(t_j), \quad j = 0, 1, 2, \ldots$, be a control, which generates the trajectory $x(0)$, $x(t_1), x(t_2), \ldots, x(t_k), \ldots$. For this control we define the mean integral-time cost by a trajectory

$$F_{x_0}\,(u(t)) = \lim_{k\to\infty} \frac{\sum_{j=1}^{k-1} c_{t_j}\left(x(t_j),\ g_{t_j}\left(x(t_j),\ u(t_j)\right)\right)}{\sum_{j=0}^{k-1} \tau_{x(t_j)}(u(t_j))}$$

where $c_{t_j}\left(x(t_j),\ g_{t_j}\left(x(t_j),\ u(t_j)\right)\right) = c_{t_j}\left(x(t_j),\ x(t_{j+1})\right)$ represents the cost of the system $\mathbb{L}$ to pass from the state $x(t_j)$ to the state $x(t_{j+1})$ at the stage $[j,\ j+1]$.

We consider the problem of finding the time-moments $t = 0, t_1, t_2, \ldots, t_{k-1}, \ldots$ and the vectors of control parameters $u(0), u(t_1), u(t_2), \ldots, u(t_{k-1}), \ldots$ which satisfy the conditions mentioned above and minimize the functional $F_{x_0}\,(u(t))$.

In the case of $\tau_{x(t)}(u(t)) \equiv 1$ for every t and $x(t)$ this problem becomes the control problem with unit time of states' transitions. The problem of determining the stationary control with unit time of states' transitions has been studied in [5, 53, 65, 73, 117]. In the mentioned papers it is assumed that $U_t(x(t))$, g_t and c_t do not depend on t, i.e., $g_t = g$, $c_t = c$ and $U_t(x) = U(x)$ for $t = 0, 1, 2, \ldots$. Richard Bellman showed in [5] that for the stationary case of the problem with unit time of states' transitions there exists an optimal stationary control $u^*(0), u^*(1), \ldots, u^*(t), \ldots$

such that

$$\lim_{k\to\infty} \frac{\sum_{t=0}^{k-1} c\big(x(t),\ g\left(x(t), u^*(t)\right)\big)}{k}$$

$$= \inf_{u(t)} \lim_{k\to\infty} \frac{\sum_{t=0}^{k-1} c\big(x(t),\ g\left(x(t), u(t)\right)\big)}{k} = \lambda < \infty.$$

Furthermore in [65, 117] it is shown that the stationary case of the problem can be reduced to the problem of finding the optimal mean cost cycle in a graph of states' transitions of a dynamical system. Based on these results in [18, 53, 73, 117] polynomial-time algorithms for finding the optimal stationary control are proposed. This variant of the problem can be solved by using the linear programming problem (2.18), (2.19) from Sect. 2.2.4.

Below we extend the results mentioned above to the general stationary case of the problem with arbitrary transit-time functions τ_x. We show that this problem can be formulated as the problem of determining the optimal mean cost cycles in the graph of states' transitions of the dynamical system for an arbitrary transition-time function on the edges.

For the discounted control problem with varying time of states' transitions the dynamics is determined in the same way as for the problem above; but the objective function which has to be minimized is defined as follows:

$$\widehat{F}_{x_0}(u(t)) = \sum_{j=0}^{\infty} \gamma^{t_j} c\big(x(t_j),\ g\left(x(t_j), u(t_j)\right)\big),$$

where $\gamma,\ 0 < \gamma < 1$, is a given discounted factor.

2.9.2 *A Linear Programming Approach for the Problem with Arbitrary Transition Costs*

We consider the stationary case of the deterministic transition control problem, i.e., when $g_t,\ c_t,\ U_t(x(t)),\ u(t)$ do not depend on t and the transition function $\tau_{x(t)}$ depends only on the state x and on the control u_x in the state x. So, $g_t = g,\ c_t = c,$ $U_t(x) = U(x),\ \tau_{x(t)} = \tau(x, u_x)$ for $u(t) = u_x \in U(x),\ \forall x \in X,\quad t = 0, 1, 2, \ldots$.

In this case it is convenient to study the problem on a network where the dynamics of the system is described by the graph of states' transitions $G = (X, E)$. An arbitrary vertex x of G corresponds to a state $x \in X$ and an arbitrary directed edge $e = (x, y) \in E$ expresses the possibility of the system $\mathbb{L}$ to pass from the state $x(t)$ to the state $x(t + \tau_e)$, where τ_e is the time of the system's transition from the state x to the state y through the edge $e = (x, y)$. So, on the edge set E it is defined the function

$\tau : E \to \mathbb{R}^+$ which associates to each edge a positive number τ_e which means that if the system $\mathbb{L}$ at the moment of time t is in the state $x = x(t)$ then the system can reach the state y at the moment of time $t + \tau_e$ if it passes through the edge $e = (x, y)$, i.e., $y = x(t + \tau_e)$. In addition, on the edge set E it is defined the cost function $c : E \to \mathbb{R}$, which associates to each edge the cost c_e of the system's transition from the state $x = x(t)$ to the state $y = x(t + \tau_e)$ for an arbitrary discrete moment of time t. So, finally we have that to each edge $e = (x, y) \in E$ the cost c_e and the transition time τ_e from x to y are associated.

In G an arbitrary edge $e = (x, y)$ corresponds to a control in the initial problem and the set of edges $E(x) = \{e = (x, y) \,|\, (x, y) \in E\}$ originating in the vertex x corresponds to the feasible set $U(x)$ of the vectors of control parameters in the state x. The transition time function τ in G is induced by the transition time function τ_x for the stationary control problem.

It is easy to observe that the infinite horizon control problem with a varying time of states' transitions of the system on G can be regarded as the problem of finding in G the minimal mean cost cycle C_G^* that can be reached from the vertex x_0 where the vertex x_0 corresponds to the starting state $x_0 = x(0)$ of the dynamical system $\mathbb{L}$. Indeed, a stationary control in G corresponds to a fixed transition from a vertex $x \in X$ to another vertex $y \in X$ through a directed edge $e = (x, y)$ in G. Such a strategy of states' transitions of the dynamical system in G generates a trajectory which leads to a directed cycle C_G with the set of edges $E(C_G)$. Therefore, the considered stationary control problem on G is reduced to the problem of finding the minimal mean cost cycle that can be reached from x_0, where in G to each directed edge $e = (x, y) \in E$ the cost c_e and the transition time τ_e of the system's transition from the state $x = x(t)$ to the state $y = x(t + \tau_e)$ are associated.

If the minimal mean cost cycle C_G^* in G is known then the stationary optimal control for our problem can be found by the following way: In G we fix an arbitrary simple directed path $P(x_0, x_k)$ with the set of edges $E(P(x_0, x_k))$ which connects the vertex x_0 with the cycle C_G^*. After that for an arbitrary state $x \in X$ we choose a stationary control which corresponds to a unique directed edge $e = (x, y) \in E(P(x_0, x_k)) \cup E(C^*)$. For such a stationary control the following equality holds:

$$\inf_{u(t)} \lim_{k \to \infty} \frac{\sum_{j=0}^{k-1} c\big(x(t_0),\ g\left(x(t_j), u(t_j)\right)\big)}{\sum_{j=0}^{k-1} \tau_x(u(t_j))} = \frac{\sum_{e \in E(\mathbb{C}^*)} c_e}{\sum_{e \in E(\mathbb{C}^*)} \tau_e}.$$

Note that the condition $U(x) \neq \emptyset,\ \forall x \in X$, for the stationary case of the control problem means that in G each vertex x contains at least one leaving directed edge $e = (x, y)$. We will assume that in G every vertex $x \in X$ is attainable from x_0; otherwise we can delete vertices from X for which there are no directed paths $P(x_0, x)$ from x_0 to x. Moreover, without loss of generality, we may consider that G is a strongly connected graph. Then the problem of finding the optimal stationary control for the problem from Sect. 2.2.4 can be formulated as combinatorial optimization problem on G in which it is necessary to find a directed cycle C_G^* such that

$$\frac{\sum\limits_{e\in E(C_G^*)} c_e}{\sum\limits_{e\in E(C_G^*)} \tau_e} = \min_{C_G} \frac{\sum\limits_{e\in E(C_G)} c_e}{\sum\limits_{e\in E(C_G)} \tau_e}.$$

The problem of determining the minimal mean cost cycle in a double weighted directed graph has been studied in [19, 53, 63, 117]. In the cited works algorithms based on linear programming and parametrical methods are proposed. For the problem with a unit time of states' transitions in [53] a strongly polynomial time algorithm is proposed.

In the following we describe an approach which is based on linear programming. We can see that such an approach may be used for solving a more general class of problems, as example, for the multi-criterion version of minimal mean cost cycle problems [82].

We consider the following linear programming problem:
Minimize

$$z = \sum_{e\in E} c_e \alpha_e \tag{2.147}$$

subject to

$$\begin{cases} \sum\limits_{e\in E^+(x)} \alpha_e - \sum\limits_{e\in E^-(x)} \alpha_e = 0, \quad \forall x \in X; \\ \sum\limits_{e\in E} \tau_e \alpha_e = 1; \\ \alpha_e \geq 0, \quad \forall e \in E. \end{cases} \tag{2.148}$$

where $E^+(x) = \{e = (x, y) \in E \mid y \in X\}$, $E^-(x) = \{e = (y, x) \in E \mid y \in X\}$.

The following lemma holds.

Lemma 2.68 *Let* $\alpha = \left(\alpha_{e_1}, \alpha_{e_2}, \ldots, \alpha_{e_m}\right)$ *be a feasible solution of the system* (2.148) and $G_\alpha = (X_\alpha, E_\alpha)$ be the subgraph of G, generated by the set of edges $E_\alpha = \{e_i \in E \mid \alpha_{e_i} > 0\}$. Then an arbitrary extreme point $\alpha^0 = \left(\alpha^0_{e_1}, \alpha^0_{e_2}, \ldots, \alpha^0_{e_m}\right)$ of the polyhedron set determined by (2.148) corresponds to a subgraph $G_{\alpha^0} = (X_{\alpha^0}, E_{\alpha^0})$ which has the structure of a simple directed cycle and vice versa, i.e., if $G_{\alpha^0} = (X_{\alpha^0}, E_{\alpha^0})$ is a simple directed cycle in G then the solution $\alpha^0 = \left(\alpha^0_{e_1}, \alpha^0_{e_2}, \ldots, \alpha^0_{e_m}\right)$ with

$$\alpha^o_{e_i} = \begin{cases} \dfrac{1}{\sum_{e\in E_{\alpha^0}} \tau_e}, & \text{if } e_i \in E_{\alpha^0}; \\ 0, & \text{if } e_i \notin E_{\alpha^0} \end{cases}$$

corresponds to an extreme point of the set of solutions (2.148).

Proof Let $\alpha = \left(\alpha_{e_1}, \alpha_{e_2}, \ldots, \alpha_{e_m}\right)$ be an arbitrary feasible solution of the system (2.148). Then it is easy to observe that $G_\alpha = (X_\alpha, E_\alpha)$ contains at least one directed cycle. Indeed, for an arbitrary $x \in X_\alpha$ there exist at least one leaving edge $e' = (x, y) \in E_\alpha$ and at least one entering edge $e'' = (z, x) \in E_\alpha$; otherwise α does not satisfy condition (2.148).

Let us show that if G_α is not a simple directed cycle then α does not represent an extreme point of the set of solutions of the system (2.148). If G_α has not the structure of a simple directed cycle then it contains a simple directed cycle C with the set of edges $E(C_G) \subset E_\alpha$, i.e., $m' = |E(C_G)| < m$. Without loss of generality we may consider that $E(\mathbb{C}) = \{e_1, e_2, \ldots, e_{m'}\}$. Fix an arbitrary value θ such that $0 < \theta < \min_{e_i \in E(\mathbb{C})} \alpha_{e_i}$ and consider the following two solutions:

$$\alpha^1 = \frac{1}{1 - \theta \sum_{i=1}^{m} \tau_{e_i}} \left(\alpha_{e_1} - \theta,\ \alpha_{e_2} - \theta,\ \ldots,\ \alpha_{e_{m'}} - \theta,\ \alpha_{e_{m'+1}},\ \ldots,\ \alpha_{e_m}\right);$$

$$\alpha^2 = \frac{1}{\theta \sum_{i=1}^{m} \tau_{e_i}} \big(\underbrace{\theta, \theta, \ldots, \theta}_{m'}, 0, 0, \ldots, 0\big).$$

It is easy to check that α^1 and α^2 satisfy the condition (2.148), i.e., α^1 and α^2 are feasible solutions of the problem (2.147), (2.148). If we chose θ such that $0 < \theta \sum_{i=1}^{m'} \tau_{e_i} < 1$ then we obtain that α can be represented as a convex combination of feasible solutions α^1 and α^2, i.e.,

$$\alpha = \left(1 - \theta \sum_{i=1}^{m'} \tau_e\right) \alpha^1 + \left(\theta \sum_{i=1}^{m'} \tau_e\right) \alpha^2. \tag{2.149}$$

So, α is not an extreme point of the set of solutions (2.148). If G_α represents a simple directed cycle then the representation (2.149) is not possible, i.e., the second part of Lemma 2.68 holds. □

Using Lemma 2.68 we can prove the following result.

Theorem 2.69 *The optimal basic solution* $\alpha^* = \left(\alpha^*_{e_1}, \alpha^*_{e_2}, \ldots, \alpha^*_{e_m}\right)$ *of problem* (2.147), (2.148) *corresponds to a minimal mean cycle* $C^*_G = G_{\alpha^*}$ *in* G, *i.e.,*

$$\alpha^*(e_i) = \begin{cases} \dfrac{1}{\sum_{e \in E(\mathbb{C}^*)} \tau_e}, & \text{if } e \in E(\mathbb{C}^*); \\ 0, & \text{if } e \notin E(\mathbb{C}^*), \end{cases}$$

where $E(C^*_G)$ *is the set of edges of a directed cycle* C^*_G.

Proof According to Lemma 2.68 an arbitrary extreme point α^0 of the set of solutions of system (2.148) corresponds in G to the subgraph $G_{\alpha^0} = (X^0_\alpha, E^0_\alpha)$ which has the

structure of a directed cycle. Taking into account that the optimal solution of problem (2.147), (2.148) is attained in an extreme point we obtain the proof of the theorem.□

The linear programming problem (2.147), (2.148) allows us to find the minimal mean cycle in the graph G with positive values $\tau_e = \tau_{x,y}$, for $e = (x, y) \in E$. More efficient algorithms for solving the problem can be obtained using the dual problem (2.147), (2.148).

Theorem 2.70 *If G is a strongly connected directed graph then there exists a function $\varepsilon : X \to \mathbb{R}$ and the value λ such that:*

(a) $\varepsilon_y - \varepsilon_x + c_{x,y} \geq \tau_{x,y} \cdot \lambda, \quad \forall (x, y) \in E;$
(b) $\min\limits_{y \in O^-(x)} \{\varepsilon_y - \varepsilon_x + c_{x,y} - \tau_{x,y}\lambda\} = 0, \quad \forall x \in X;$
(c) *an arbitrary cycle $\mathbb{C}^*$ of the subgraph $G^0 = (X, E^0)$ of G, generated by edges $(x, y) \in E$ for which $\varepsilon_y - \varepsilon_x + c_{x,y} - \tau_{x,y} \cdot \lambda = 0$ determines a minimal mean cycle in G.*

Proof We consider the dual problem for (2.147), (2.148):
Maximize

$$W = \lambda$$

subject to

$$\varepsilon_x - \varepsilon_y + \tau_{x,y}\lambda \leq c_{x,y}, \quad \forall (x, y) \in E.$$

If p is the optimal value of the problem then by using duality properties of the solution of the problem we obtain (a), (b) and (c). □

Based on results described above we can make the following conclusions.

1. If $\lambda = 0$ then the values $\varepsilon_x, \ x \in X$ can be treated as the cost of minimal paths from vertices $x \in X$ to a vertex x_f which belongs to the minimal mean cycle C^*_G (with $\lambda = 0$) in the graph G with given costs c_e of edges $e \in E$. So, if x_f is known then the cycle C^*_G can be found in the following way. We construct the tree of minimum cost directed paths from $x \in X$ to x_f and determine the values $\varepsilon_x, \ \forall x \in X$. Then in G we make a transformation of the costs $c'_{x,y} = \varepsilon_y - \varepsilon_x + c_{x,y}$ for $(x, y) \in E$ and find the subgraph $G^0 = (X, E^0)$ generated by edges (x, y) with $c'_{x,y} = 0$. After that we fix in G^0 a cycle C^* with zero cost of the edges. If the vertex x_f is not known then we have to construct the tree of minimal cost paths which respect to each $x_f \in X$. So, in this case with respect to each tree we find the subgraph $G^0 = (X, E^0)$. Then at least for one of such a subgraph we find a cycle C^*_G with zero cost ($c'_{x,y} = 0$) of the edges.
2. If $\lambda \neq 0$ and λ is known then the minimal mean cost cycle C^* can be found by the following way. In G we change the costs $c_{x,y}$ of edges $(x, y) \in E$ by $c_{x,y} - \tau_{x,y}\lambda$ and after that solve the problem with the new costs according to point 1.

3. If $\lambda \neq 0$ and it is not known then we find it using the bisection method on the segment $[h_0^1, h_0^2]$ where $h^1 = \min_{e\in E} c_e$, $h^2 = \max_{e\in E} c_e$. At each step k of the method we find the midpoint $\lambda_k = \left(h_k^1 + h_k^2\right)/2$ of the segment $[h_k^1, h_k^2]$ and check if in G with the cost $c_{x,y}^k - \tau_{x,y}\lambda_k$ there exists the cycle with negative cost. If at a given step there exists the cycle with negative cost then we fix $h_{k+1}^1 = h_k^1$, $h_{k+1}^2 = \lambda_k$; otherwise we put $h_{k+1}^1 = \lambda_k$, $h_{k+1}^1 = h_k^1$. In such a way we find λ with a given precision. After that the exact value of λ can be found from λ_k using a special roundoff procedure from [58].

The algorithm described above allows us to determine the solution of the problem in the case if $\tau_e \geq 0, \forall e \in E$. In general this problem can be considered for arbitrary τ_e and c_e. In this case we may use the following fractional linear programming problem: Minimize

$$z = \frac{\sum_{e\in E} c_e\alpha_e}{\sum_{e\in E} \tau_e\alpha_e} \tag{2.150}$$

subject to

$$\begin{cases} \sum\limits_{e\in E^+(x)} \alpha_e - \sum\limits_{e\in E^-(x)} \alpha_e = 0, \ \forall x \in X; \\ \sum\limits_{e\in E} \alpha_e = 1; \\ \alpha_e \geq 0, \ e \in E, \end{cases} \tag{2.151}$$

where $E^-(x) = \{e = (y, x) \in E \mid y \in X\}$; $E^+(x) = \{e = (x, y) \in E \mid y \in X\}$.

Of course, this model is valid if on the set of solutions of system (2.151) it holds $\sum_{e\in E} \tau_e\alpha_e \neq 0$. In a similar way as for the linear programming problem here we can show that an arbitrary optimal basic solution of the problem (2.150), (2.151) corresponds to an optimal mean directed cycle in G.

Let $\alpha = \left(\alpha_{e_1}, \alpha_{e_2}, \ldots, \alpha_{e_{|E|}}\right)$ be an arbitrary feasible solution of system (2.151) and denote by $G_\alpha = (X_\alpha, E_\alpha)$ the subgraph of G generated by the set of edges $E_\alpha = \{e \in E \mid \alpha_e > 0\}$. In [73] it is shown that an arbitrary extreme point $\alpha^0 = \left(\alpha_{e_1}^0, \alpha_{e_2}^0, \ldots, \alpha_{e_{|E|}}^0\right)$ of the set of solutions of system (2.151) corresponds to a subgraph $G_{\alpha^0} = (X_{\alpha^0}, E_{\alpha^0})$ which has the structure of an elementary directed cycle. Taking into account that for the problem (2.150), (2.151) there exists an optimal solution $\alpha^* = \left(\alpha_{e_1}^*, \alpha_{e_2}^*, \ldots, \alpha_{e_{|E|}}^*\right)$ which corresponds to an extreme point of the set of solutions (2.151) we obtain that

$$\max z = \frac{\sum_{e\in E_{\alpha^*}} c_e\alpha_e^*}{\sum_{e\in E_{\alpha^*}} \tau_e\alpha_e^*}$$

and the set of edges E_{α^*} generates a directed cycle G_{α^*} for which $\alpha_e^* = 1/|E_{\alpha^*}|$, $\forall e \in E_{\alpha^*}$. Therefore,

$$\max z = \frac{\sum_{e \in E_{\alpha^*}} c_e}{\sum_{e \in E_{\alpha^*}} \tau_e}.$$

So, an optimal solution of problem (2.150), (2.151) corresponds to the minimal mean cost cycle in the directed graph of states' transitions of the dynamical system.

This means that the fractional linear programming problem (2.150), (2.151) can be used for determining the optimal solution of the problem in the general case.

2.9.3 *Reduction of the Problem to the Case with Unit Time of States' Transitions*

As we have shown the deterministic control problem with an average cost criterion on the network can be solved for an arbitrary transition-time function using a linear programming problem (2.147), (2.148) or a linear fractional programming problem (2.150), (2.151). For the discounted control problem with varying time of state transitions a similar linear programming model could not be derived. However, both problems can be reduced to the corresponding cases of the problems with unit time of states' transitions of the system.

Below we describe a general scheme how to reduce the control problems with varying time of states' transitions to the case with unit time of states' transition of the system. We show that our problems can be reduced to the case with unit time of states' transitions on an auxiliary graph $G' = (X', E')$ which is obtained from $G = (X, E)$ using a special construction. This means that after such a reduction we can apply the linear programming approach described in Sect. 2.2.

Graph $G' = (X', E')$ with unit transitions on directed edges $e' \in E'$ is obtained from G where each directed edge $e = (x, y) \in E$ with corresponding transition time τ_e is changed by a sequence of directed edges

$$e'_1 = (x, x_1^e),\ e'_2 = (x_1^e, x_2^e),\ \ldots,\ e'_{\tau_e} = (x_{\tau_e - 1}^e, y).$$

This means that we represent a transition from a state $x = x(t)$ at the moment of time t to the state $y = x(t + \tau_e)$ at the moment of time $t + \tau_e$ in G in G' as the transition of a dynamical system from the state $x = x(t)$ at the time-moment t to $y = x(t + \tau_e)$ if the system makes transitions through a new fictive intermediate set of states $x'_1, x'_2, \ldots, x'_{\tau - 1}$ at the corresponding discrete moments of time

$$t + 1,\ t + 2,\ \ldots,\ t + \tau_e - 1.$$

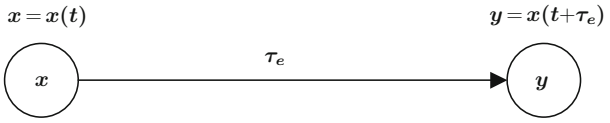

Fig. 2.14 The edge $e = (x, y)$ with the associated transition time τ_e

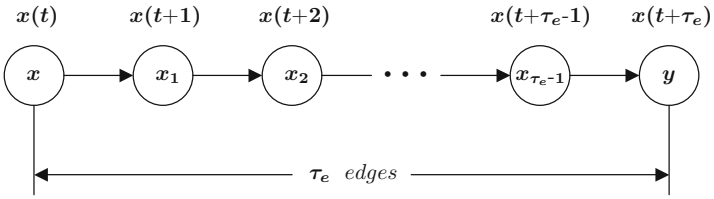

Fig. 2.15 The intermediate states for the edge $e = (x, y)$ in G'

The graphical interpretation of this construction is represented in Figs. 2.14 and 2.15. In Fig. 2.14 it is represented an arbitrary directed edge $e = (x, y)$ with the corresponding transition time τ_e in G. In Fig. 2.15 it is represented the sequence of directed edges e'_i and the intermediate states $x_1, x_2, \ldots, x_{\tau_e - 1}$ in G' that correspond to a directed edge $e = (x, y)$ in G. So, the set of of vertices X' of the graph G' consists of the set of states X and the set of intermediate states $XE = \{x_i^e \mid e \in E,\ i = 1, 2, \ldots, \tau_e\}$, i.e., $X' = X \cup XE$. Then the set of edges E' is defined as follows:

$$E' = \bigcup_{e \in E} \mathcal{E}^e, \quad \mathcal{E}^e = \{(x, x_1^e),\ (x_1^e, x_2^e),\ \ldots,\ (x_{\tau_e - 1}^e, y) \mid e = (x, y) \in E\}.$$

We define the cost function $c' : E' \to \mathbb{R}$ in the following way:

$$c'_{x,x^e} = c_{x,y}, \quad \text{if} \quad e = (x, y) \in E;$$
$$c'_{x_1^e, x_2^e} = c_{x_2^e, x_3^e} = \cdots = c_{x_{\tau_e - 1}, y} = 0.$$

It is evident that between the set of stationary strategies

$$s : x \to y \in X + (x) \quad \text{for} \quad x \in X$$

and the set of stationary strategies

$$s' : x' \to y' \in X'^{+}(x') \quad \text{for} \quad x' \in X'$$

there exists a bijective mapping such that the corresponding average and discounted costs on G and on C' are the same. So, if s'^* is the optimal stationary strategy of the problem with unit transitions on G' then the optimal stationary strategy s^* on

G is determined by fixing $s^*(x) = y$ if $s'^*(x) = x_1^e$, where $e = (x, y)$. For the stochastic versions of the control problem on $G = (X, E)$ the construction of the auxiliary graph is similar. Here we should take into account that the set of vertices (states) X are divided into two disjoint subsets X_C and X_N where X_C correspond to the set of controllable states and X_N corresponds to the set of uncontrollable states. Moreover, the probability function $p : E_N \to [0, 1]$ on the set $E_N = \{e = (x, y) \in E \mid x \in X_N\}$ is defined such that $\sum_{y \in X^+(x)} p_{x,y} = 1$. The graph $G' = (X', E')$ in the case of stochastic control problems is constructed in the same way as above. Here we have only to precise how to define the sets X'_C, X'_N and the probability function p' on the set $E'_N = \{e' = (x', y') \in E' \mid x' \in X'_N\}$ in G'. To obtain a bijective mapping between the stationary strategies of the problems in the initial graph G and the stationary strategies of the problem in the auxiliary graph it is necessary to take $X'_C = X_C$, $X'_N = X' \backslash X_C$ and to define the probability function $p' : E' \to [0, 1]$ as follows:

$$p'_{x',y'} = \begin{cases} p_{x,y}, & \text{if } x' = x,\ x' \in X_N \subset X'_N \quad \text{and} \quad y' = x'_1; \\ 0, & \text{if } x' \in X'_N \backslash X_N. \end{cases}$$

The cost function on G' for the corresponding auxiliary stochastic control problems is defined in the same way as for deterministic problems.

In the following we extend the approach described above to Semi-Markov decision problems, which is valid for the stochastic control problem in its general form.

2.10 Determining the Optimal Strategies for Semi-Markov Decision Problems

The average and discounted Markov decision problems can be extended to Semi-Markov Decision Processes [113–115, 134, 140, 141]. A Semi-Markov decision process is determined by a finite state space X, a finite set of actions A, a nonnegative real function

$$p : A \times X \times X \times \{1, 2, \ldots, \bar{t}\} \to [0, 1]$$

that satisfies the condition

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

and the cost function

$$c : A \times X \times X \times \{1, 2, \ldots, \bar{t}\} \to \mathbb{R}.$$

Here the function p for a fixed action $a \in A$, arbitrary $x, y \in X$ and a fixed $\tau \in \{1, 2, \ldots, \bar{t}\}$ determines the probability $p^a_{x,y,\tau}$ of the system to pass from the state $x \in X$ to state y by using τ units of time. The function c for a fixed action a in the state $x \in X$, a given $y \in X$ and a fixed τ determines the cost $c^a_{x,y,\tau}$ of the system to pass from the state x to the state y using τ units of time. We define a stationary strategy s in the Semi-Markov decision process as a map

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

where $A(x)$ represents the set of actions in the state $x \in X$. An arbitrary stationary strategy s induces a Semi-Markov process with the transition probabilitis $p^s_{x,y,\tau}$ and the transition costs $c^s_{x,y,\tau}$. For this Semi-Markov process with given transition costs we can define the average cost per transition $\omega_{x_0}(s)$ and the expected total discounted cost $\sigma^{\gamma}_{x_0}(s)$ if the system starts transitions in the state x_0 at the moment of time $t = 0$. The problems of determining stationary strategies with minimal average and expected total discounted cost for Semi-Markov decision processes can be formulated and studied in a similar way as for Markov decision processes.

Using the results from Sect. 1.9 we can reduce the considered decision problems to the corresponding problems for an auxiliary Markov decision process. Indeed, for an arbitrary action $a \in A$ in a state $x \in X$, a given $y \in X$ and fixed $\tau \in \{1, 2, \ldots, \bar{t}\}$ the transition from x to y in Semi-Markov decision process we represent as a sequence of τ transitions with unit time via τ fictive intermediate states

$$\begin{gathered} x \to x_1^{a,\tau}, \\ x_1^{a,\tau} \to x_2^{a,\tau}, \ \ldots, \ x_{\tau-2}^{a,\tau} \to x_{\tau-1}^{a,\tau}, \\ x_{\tau-1}^{a,\tau} \to y, \end{gathered}$$

where the corresponding probabilities and transition costs are defined as follows:

$$\begin{aligned} p_{x,x_1^{a,\tau}} &= p^a_{x,y,\tau}; \\ p_{x_1^{a,\tau},x_2^{a,\tau}} &= p_{x_2^{a,\tau},x_3^{a,\tau}} = \cdots = p_{x_{\tau-2}^{a,\tau},x_{\tau-1}^{a,\tau}} = p_{x_{\tau-1}^{a,\tau},y} = 1; \\ c^a_{x,x_1^{a,\tau}} &= c^a_{x,y,\tau}; \\ c^a_{x_1^{a,\tau},x_2^{a,\tau}} &= c^a_{x_2^{a,\tau},x_3^{a,\tau}} = \cdots = c^a_{x_{\tau-2}^{a,\tau},x_{\tau-1}^{a,\tau}} = c^a_{x_{\tau-1}^{a,\tau},y} = 0. \end{aligned}$$

After that we consider a new Markov decision problem with a new set of states $\overline{X}' = X \cup X'$ obtained from X by adding the set of fictive intermediate states X' and new probability and cost functions defined above; the set of actions in the state $x \in X$ in the new problem are the same as for Semi-Markov decision problems and in each added fictive state there is a unique action determined by a unique transition to the next state with a probability equal to 1.

It is evident that if $\overline{s}^*$ is a optimal stationary strategy for the auxiliary decision problem then an optimal stationary strategy s^* of the Semi-Markov decision problem (with average or discounted optimization criterion) can be found in the following way:

$$s^*(x) = \overline{s}^*(x) \quad \text{for} \quad x \in X.$$

In such a way we can reduce the Semi-Markov decision problem to the corresponding auxiliary Markov decision problem.

http://www.springer.com/978-3-319-11832-1

Optimization of Stochastic Discrete Systems and Control on Complex Networks
Computational Networks
Lozovanu, D.; Pickl, S.
2015, XIX, 400 p. 54 illus., Hardcover
ISBN: 978-3-319-11832-1