We consider a Markovian decision process with a nonhomogeneous transition function satisfying a periodicity condition. An optimization method is proposed which computes the optimal periodic strategy for an unbounded time interval.

INTRODUCTION

Optimal decision methods in stochastic systems described by a controlled Markov chain (a Markovian decision process) with a homogeneous transition function were described in [1, 2]. There is, however, a wide class of systems in which the transition function is nonhomogeneous, i.e., time dependent. In particular, this class includes systems whose output parameters essentially depend on periodically varying external conditions (e.g., gas transport systems, power supply systems, water reservoir control systems, etc.).

In this study, we develop an optimal control method on an unbounded time interval for a stochastic system with a nonhomogeneous transition function that satisfies a periodicity condition. The formal description of the controlled stochastic system relies on the Markovian decision process model.

STATEMENT OF THE PROBLEM

A Markovian decision process is described by the tuple \( \{E, Y, \sigma, q_t, w_t\} \), where \( E \) is the state set, \( Y \) is the set of controls (decisions), \( \sigma \) is the control choosing rule, \( q_t, t = 1, 2, ..., \) is the transition function that defines for each \( t \) the probability distribution on the set \( E \) for every pair \( (x, y), x \in E, y \in Y; w_t, t = 1, 2, ..., \) is the income function defined on \( E \) for each \( y \in Y \).

We assume that \( E \) and \( Y \) are finite sets, and the functions \( q_t \) and \( w_t \) satisfy the periodicity condition

\[
q_t = q_{t+T}, \quad t = 1, 2, ..., \\
w_t = w_{t+T}, \quad t = 1, 2, ..., \tag{1, 2}
\]

where \( T \) is a positive integer. The minimum \( T \) satisfying (1) and (2) is called the period.
The mapping \( \pi : E \to Y \) is called a decision function and the sequence of decision functions \( \{\pi_1, \pi_2, \ldots\} \) is the strategy \( \sigma \), which defines a rule for choosing the controls \( y \in Y \) at each time moment \( t = 1, 2, \ldots \).

The strategy \( \sigma = \{\pi_1, \pi_2, \ldots\} \) is \( T \)-periodic if \( \pi_t = \pi_{t+T}, t = 1, 2, \ldots \), where \( T \geq 1 \). The class of \( T \)-periodic strategies will be denoted by \( \Pi(T) \). If \( T = 1 \), then strategies of class \( \Pi(1) \) are called stationary. A strategy consisting of finitely many decision functions will be called truncated.

Let some period \( T \) be given. To a periodic strategy \( \sigma \in \Pi(T) \) we associate the mean total income on the time interval \([1, nT]\),

\[
\varphi_n(\sigma)(x) = \frac{1}{nT} \sum_{t=1}^{nT} \varphi_t(x, y_t),
\]

where \( x = x_1 \) is the initial state in \( E \), \( \varphi_t(x_t, y_t) \) is the income earned using the control \( y_t = \pi_t(x_t) \) in state \( x_t \in E \) at time \( t \). The expectation in (3) is evaluated over the measure generated by the transition function \( q_t(x, y_t) \), \( x, y_t \in Y, x \in E \).

The strategy \( \sigma^* \) such that \( \varphi_n(\sigma^*)(x) \geq \varphi_n(\sigma)(x) \) for all \( \sigma \in \Pi(T) \) and \( x \in E \) will be called \( \varphi_n \)-optimal.

For the strategy \( \sigma \in \Pi(T) \) we also define the mean income per unit time \( \varphi(\sigma)(x) = \lim_{n \to \infty} \varphi_n(\sigma)(x) \). The strategy \( \sigma^* \) such that \( \varphi(\sigma^*)(x) \geq \varphi(\sigma)(x) \) for all \( \sigma \in \Pi(T) \) on \( x \in E \) is called \( \varphi \)-optimal.

The problem is to find a \( \varphi_n \)-optimal or a \( \varphi \)-optimal strategy. We consider the problem of finding a \( \varphi \)-optimal strategy on the assumption that for any strategy \( \sigma \) the set \( E \) constitutes one ergodic class of states [1].

**AUXILIARY CONSTRUCTS**

To each decision function \( \pi \) we can associate the transition probability matrix \( Q(\pi) \) with the elements \( q(z|x, \pi(x)) \), \( z, x \in E \), and an income column vector \( w(\pi) \) with the components \( w(\pi)(x) = w(x, \pi(x)) \), \( x \in E \).

In what follows, the sequence \( \{a_1, a_2, \ldots, a_n\} \) is denoted in abbreviated form as \( a_1^n \).

Let \( \delta = \pi_1^T \) be the collection of decision functions corresponding to the period \( T \). We can form a periodic strategy \( \{\delta, \delta, \ldots\} \in \Pi(T) \), which we denote by \( \delta^\infty \).

**LEMMA 1.** Let \( \delta^\infty \in \Pi(T) \) and \( \delta = \pi_1^T \). Then the mean income vector \( \varphi(\delta^\infty) \) has equal components

\[
\varphi(\delta^\infty)(x) = \lim_{n \to \infty} \frac{1}{nT} \sum_{t=1}^{nT} \varphi_t(x, y_t), \quad x \in E,
\]

and is representable in the form

\[
\varphi(\delta^\infty) = Q^*(\delta) w(\delta),
\]

Here

\[
Q^*(\delta) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} Q^t(\delta), \quad Q(\delta) = \prod_{t=1}^T Q(\pi_t),
\]

\[
w(\delta) = \frac{1}{T} \sum_{k=1}^k \prod_{t=1}^T Q_{t-1}(\pi_{t-1}) w_k(\pi_k),
\]

\(Q_0(\pi)\) is the identity matrix.

**Proof.** Since \( Q(\delta) \) is a stochastic matrix for any vector \( \delta \), then \( Q^t(\delta) \) exists [2]. Denote by \( Q_t(\delta) = \Pi_{k=1}^T Q_k(\pi_k) \) the \( t \)-step transition matrix corresponding to the strategy \( \delta^\infty \). Then the component of the expected income vector at time \( t+1 \) given the initial state \( x \in E \) is \( (Q_t(\delta) w_{t+1}(\pi_{t+1}))(x) \). The mean income vector \( \varphi(\delta^\infty) \) in unit time may be written as

\[
\varphi(\delta^\infty) = \lim_{n \to \infty} \frac{1}{nT} \sum_{t=0}^{nT-1} Q_t(\delta) w_{t+1}(\pi_{t+1}),
\]

where \( Q_0(\delta) \) is the identity matrix.