ASYMPTOTICALLY OPTIMAL SYSTEMS AS A MODEL OF THE PROCESS OF LEARNING CONTROL

1966

SovietRxiv

Abstract

Full Text

UDC 621.3.078

CYBERNETICS AND CONTROL THEORY

Yu. P. LEONOV

ASYMPTOTICALLY OPTIMAL SYSTEMS AS A MODEL OF THE PROCESS OF LEARNING CONTROL

(Presented by Academician B. N. Petrov on June 29, 1965)

Let there be an object with a controlled coordinate $\{Y_n\}$ and a control signal $\{X_n\}$—both signals being discrete functions of time. The object is statistical, so that when $\{X_n\} = \{x_n\}$ (where $\{x_n\}$ is a realization of the random function $\{X_n\}$) the output signal is a random function. It is necessary to control the object in such a way that the control system is optimal in the sense of some quality criterion. The latter may have the form

$$J = MQ_n(Y_n, X_n, \alpha),$$

where $\mathbf{M}$ is mathematical expectation and Q is some function of the random arguments Y, X and the nonrandom quantity α. Let $\mathbf{M}\{Q/x\}$ be the conditional mathematical expectation of the quantity Q under the condition $X = x$. Then the quantity J, depending on the probability density $\varphi(x)$ of the quantity X, also has an extremum at $x = x^*$.

Now let $X_1, \ldots, X_n$ be a sequence of random variables converging, in the probabilistic sense, to x^*. For example, convergence may be understood in the mean-square sense:

$$\lim_{n\to\infty} X_n = x^*.$$

The sequence $\{X_n\}$ determines the signal controlling the object. Such control is optimal in the asymptotic sense:

$$\lim_{n\to\infty} MQ_n(Y_n, X_n, \alpha) = MQ(Y, X_n, \alpha) = \operatorname{extr}. \tag{1}$$

To find the sequence $\{X_n\}$, one may use the recurrence relation

$$x_{n+1} = x_n + \frac{a_n}{b_n}[Z_n^+ - Z_n^-] \quad (n = 1, 2, ...), \tag{2}$$

where x_1 is chosen arbitrarily; a_n and b_n are sequences of nonrandom numbers satisfying the conditions

$$a_n \to 0, \qquad b_n \to 0, \qquad \sum_1^\infty a_n = \infty, \qquad \sum_1^\infty a_n^2 < \infty,$$

$$\sum_1^\infty a_n b_n < \infty, \qquad \sum_1^\infty \left(\frac{a_n}{b_n}\right)^2 < \infty;$$

$$Z_n = Q(Y_n, X_n, \alpha);$$

Z_n^+ is obtained for the input signal $x_n + b_n$, and Z_n^- for the input signal $x_n - b_n$.

As shown in (1), the sequence $\{X_n\}$ from (2) converges in probability or in the mean square to the value x^*, giving an extremum

regression function:

$$R(x) = \mathbf{M}\{Q/x\}.$$

Consequently, (2) is the equation of an asymptotically optimal system with feedback.

If $Q(Y, X, \alpha)$ is a quadratic form in $Y - \alpha$, then one may use the simpler recurrent procedure (2)

$$x_{n+1} = x_n + \frac{A}{n}(y_n - \alpha) \qquad (n = 1, 2, ...),$$

where x_1 and A are arbitrary numbers. In this case X_n converges in probability to x^*, where x^* is the root of the equation

$$R_1(x^*) = \mathbf{M}(Y/x^*) = \alpha$$

and coincides with the x^* that gives the minimum of the function $R(x)$.

The possibility of control according to algorithm (2) can be interpreted as follows. Suppose there is some object with random disturbances. The equations of the object are unknown, as are the statistical characteristics of the disturbances. It is necessary to control the object so that the criterion J attains an extremum. At the same time, the analytic form of the dependence $Q(Y, X, \alpha)$ on Y, X, α and on the random disturbance is unknown, and there is only the possibility of

measuring the criterion at each moment of time. Then one can construct a system which, as time increases, $n \to \infty$, tends to the optimal one. Such a system is naturally called asymptotically optimal. **This result can be interpreted as the training of a system for optimal control.**

Indeed, initially nothing is known about the system,* and we cannot achieve optimal control. However, with the passage of time the system approaches the optimal one and becomes optimal as $n \to \infty$.

The learning process is formalized by means of the nonhomogeneous Markov chain (2).

Institute of Automation and Telemechanics
(Technical Cybernetics)

Received
18 VI 1965

CITED LITERATURE

[1] J. Kiefer, J. Wolfowitz, Ann. Math. Stat., **23**, No. 3, 462 (1952). [2] H. Robbins, S. Monro, Ann. Math. Stat., **22**, No. 3, 400 (1951). [3] E. T. Gladyshev, *Theory of Probability and Its Applications*, **10**, No. 2 (1965).

---

* The convergence of x_n according to (2) to the value x_0 at which the extremum of J occurs is achieved under very broad conditions (2, 3). Therefore one may consider that controllability is present when almost nothing is known about the equation of the object and the acting random disturbances.

Note: Figure translations are in progress. See original paper for figures.