Aplikace matematiky

Ivan Brůha

Learning extremal regulator implementation by a stochastic automaton and stochastic approximation theory

Aplikace matematiky, Vol. 25 (1980), No. 5, 315–323

Persistent URL: http://dml.cz/dmlcz/103867

LEARNING EXTREMAL REGULATOR IMPLEMENTATION BY A STOCHASTIC AUTOMATON AND STOCHASTIC APPROXIMATION THEORY

IVAN BRŮHA

(Received April 4, 1975)

1. INTRODUCTION

There exist many different approaches to the investigation of the characteristics of learning systems. These approaches use different branches of mathematics, e.g. gradient methods, mathematical programming, statistical decision, potential functions etc. These various approaches being taken as starting points, different results have been obtained. Some of them are too complicated and inapplicable in practice while others, on the contrary, do not match the results of practical experiments.

This paper presents one example of the modelling of learning systems by means of a stochastic automaton (abbreviation SA, cf. [1]). A basis for studying automata as learning systems can be found in [2], [3], [4], where the behaviour of automata in a random environment is studied. For a brief definition of SA see Appendix 1.

2. DESCRIPTION OF THE MODEL

Extremal control is used in the case when the extreme value of the controlled variable x is required to be at the output of the controlled system and at the same time the static time-dependent characteristic of the controlled system is unknown.

In this paper, the extremal regulator will be modelled by SA with variable structure, i.e. by SA whose transition matrix is time-dependent according to an algorithm

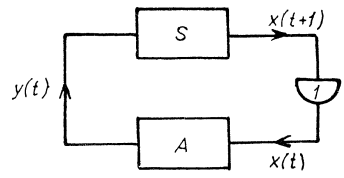

Fig. 1. The general block diagram of the model.

with unknown parameters. A block diagram is in Fig. 1, where S means the controlled system, A is SA as an extremal regulator. Let the action variable y lie in an interval $\mathscr{Y} = \langle y_{\min}, y_{\max} \rangle$ and the controlled variable x in an interval $(0, x_{\max}\rangle$.

We shall assume the existence of noise which can influence the value of the action variable y. Let us suppose that the maximum amplitude of the noise is $\pm \Delta y/2$.

First of all we shall find the minimum of the variable x. A general way of finding this minimum, while considering the noise of the regulator, would be rather difficult. We shall therefore simplify the problem by the assumption that the action variable y changes its value within intervals of the length Δy. The interval $\mathscr{Y}$ will be then divided into

$$l = \frac{y_{\max} - y_{\min}}{\Delta y} \tag{1}$$

intervals

$$\mathscr{Y}_i = \langle y_{\min} + (i-1)\,\Delta y,\ y_{\min} + i\,\Delta y \rangle\,, \quad i = 1, \ldots, l \tag{2}$$

each of the length Δy. The center of the interval $\mathscr{Y}_i$ is given by

$$\mathring{y}_i = y_{\min} + (i - \tfrac{1}{2})\,\Delta y\,, \quad i = 1, \ldots, l\,. \tag{3}$$

Let us attach a state q_i to each interval $\mathscr{Y}_i$, $i = 1, \ldots, l$. If SA is in the state q_i then a value $y \in \mathscr{Y}_i$ with uniform distribution $R(\mathring{y}_i, \Delta y)$ will be the output signal of SA.

To change its structure, SA requires information about the values of variables $x = S(y)$ for $y \in \mathscr{Y}_i$, $i = 1, \ldots, l$. The following variables are therefore introduced:

$$u_i = \frac{1}{x^n} = \frac{1}{[S(y)]^n} \quad \text{for} \quad y \in \mathscr{Y}_i\,, \quad i = 1, \ldots, l\,, \quad n \text{ is an integer}\,. \tag{4}$$

The mean value of u_i within the interval $\mathscr{Y}_i$ is

$$U_i = \frac{1}{\Delta y}\int_{\mathscr{Y}_i} u_i \,\mathrm{d}y = \frac{1}{\Delta y}\int_{\mathscr{Y}_i} \frac{1}{[S(y)]^n}\,\mathrm{d}y\,, \quad i = 1, \ldots, l\,. \tag{5}$$

It is obvious that the mean value of a random variable u_i is

$$\mathsf{E}u_i = U_i\,, \quad i = 1, \ldots, l\,.$$

Thus the problem of extremal control consists in finding such $i^* \in \{1, \ldots, l\}$ that

$$U_{i^*} = \max_{j=1,\ldots,l} U_j\,. \tag{6}$$

However, the main difficulty of solving this problem is the unknown shape of the function $S(y)$, $y \in \mathscr{Y}_i$. The only information about this shape can be obtained during single steps of the algorithm actions. We shall therefore determine estimates of $\mathsf{E}u_i$, $i = 1, \ldots, l$ by means of the so-called *linear reinforcement algorithm*; the estimate of $\mathsf{E}u_i$ for the time (step) t will be denoted $\hat{\mathsf{E}}^{(t)}u_i$.

SA as the extremal regulator works as follows: let the state distribution of SA in step t $(t = 0, 1, \ldots)$ be

$$\pi^{(t)} = [\pi_1^{(t)}, \ldots, \pi_l^{(t)}] .$$

The automaton is transferred *randomly* into another state $q(t) = q_r$ according to this distribution. The output of the automaton is then $y(t) \in \mathcal{Y}_r$ in view of the uniform distribution $R(\mathring{y}_r, \Delta y)$.

The action variable $y(t)$ on the input of the system determines the value of the controlled variable for the next step:

$$x(t + 1) = S(y(t)) \tag{7}$$

where S is the unknown static characteristic of the controlled system.

According to (4) the value of u_r for the step $t + 1$ is given by

$$u_r(t + 1) = \frac{1}{[x(t + 1)]^n}$$

and the linear reinforcement algorithm gives estimates of $\mathsf{E}u_i$ in the form $(0 < \alpha < 1)$

$$\hat{\mathsf{E}}^{(t+1)}u_r = \alpha\hat{\mathsf{E}}^{(t)}u_r + (1 - \alpha)\, u_r(t + 1) , \tag{8}$$

$$\hat{\mathsf{E}}^{(t+1)}u_j = \hat{\mathsf{E}}^{(t)}u_j , \quad j \neq r , \quad j = 1, \ldots, l .$$

The state distribution for the step $t + 1$ will be

$$\pi_i^{(t+1)} = \frac{\hat{\mathsf{E}}^{(t+1)}u_i}{\sum_{j=1}^{l} \hat{\mathsf{E}}^{(t+1)}u_j} , \quad i = 1, \ldots, l . \tag{9}$$

According to $\pi^{(t+1)}$ SA will be transferred randomly into the state $q(t + 1)$ etc. So SA works as a learning extremal regulator.

Comment. It is suitable to evaluate the average of the measured values of the controlled variable for each step

$$x_{\text{aver}}(t) = \frac{1}{t} \sum_{\tau=1}^{t} x(\tau) , \quad t = 1, 2, \ldots . \tag{10}$$

The following recurrent formula offers the most convenient way for the calculation:

$$x_{\text{aver}}(t + 1) = \frac{1}{t + 1} (t x_{\text{aver}}(t) + x(t + 1)) , \quad t = 0, 1, \ldots . \tag{11}$$

Theorem 1. *Let us assume that the formulas* (8) *are valid,* $0 < \alpha < 1$, $\mathsf{E}\hat{\mathsf{E}}^{(0)}u_i = U_i$, $\mathsf{D}u_i(t) \leqq D_{\max}$, $t = 1, 2, \ldots$ *and* $\mathsf{D}\hat{\mathsf{E}}^{(0)}u_i \leqq D_{\max}$, $i = 1, \ldots, l$. *Then*

1) $\mathsf{E}\hat{\mathsf{E}}^{(t)}u_i = U_i$, $\quad t = 0, 1, \ldots$, $\quad i = 1, \ldots, l$,

2) $(\forall \varepsilon > 0) \lim_{t \to \infty} \mathsf{P}(|\hat{\mathsf{E}}^{(t)}u_i - U_i| > \varepsilon) \leqq \frac{D_{\max}}{\varepsilon^2} \frac{1 - \alpha}{1 + \alpha}$, $\quad i = 1, \ldots, l$,

3) $$\mathrm{P}\left(\lim_{t\to\infty}\frac{1}{t+1}\sum_{\tau=0}^{t}\hat{\mathrm{E}}^{(\tau)}u_i = U_i\right) = 1\,,$$

$$\lim_{t\to\infty}\mathrm{E}\left(\frac{1}{t+1}\sum_{\tau=0}^{t}\hat{\mathrm{E}}^{(\tau)}u_i - U_i\right)^2 = 0\,, \quad i = 1, \ldots, l\,.$$

Proof of this theorem follows from the independence of $\hat{\mathrm{E}}^{(0)}u_i$ and from the values of u_i, measured during the operation of SA, from the relationship formula $\mathrm{E}u_i(t) = U_i$ and from Theorem 3 which is presented in Appendix 2.

Consequence. According to Theorem 1 we can say that the value $\pi_{i^*}^{(\infty)}$, due to the fact that the relation (6) is valid for i^*, is the maximum value of all the values $\pi_j^{(\infty)}$, $j = 1, \ldots, l$, with the probability equal to 1.

The flow diagram of the operation of SA, the program in ALGOL and several examples are presented in [6]. For better comprehension, one example is presented in this paper. Let $y_{\min} = 0$, $y_{\max} = 5$, $\Delta y = 0{\cdot}5$, $x_{\max} = 10$, $\alpha = 0{\cdot}99$, $n = 1$ and let

$$S_1(y) = \frac{2{\cdot}5y^2 - 3{\cdot}5y + 1{\cdot}8}{y + 0{\cdot}2}$$

be valid for the first 1000 steps and then

$$S_2(y) = \frac{2{\cdot}5y^2 - 21{\cdot}5y + 46{\cdot}8}{-y + 5{\cdot}2}$$

for next 1500 steps. We can see that $l = 10$, minimum of $S_1(y)$ $(S_2(y))$ equal to 0·6 corresponds to $\mathscr{Y}_2(\mathscr{Y}_9)$. The shapes of $\pi_2^{(t)}$, $\pi_9^{(t)}$, $x_{\mathrm{aver}}(t)$ can be seen in Fig. 2.

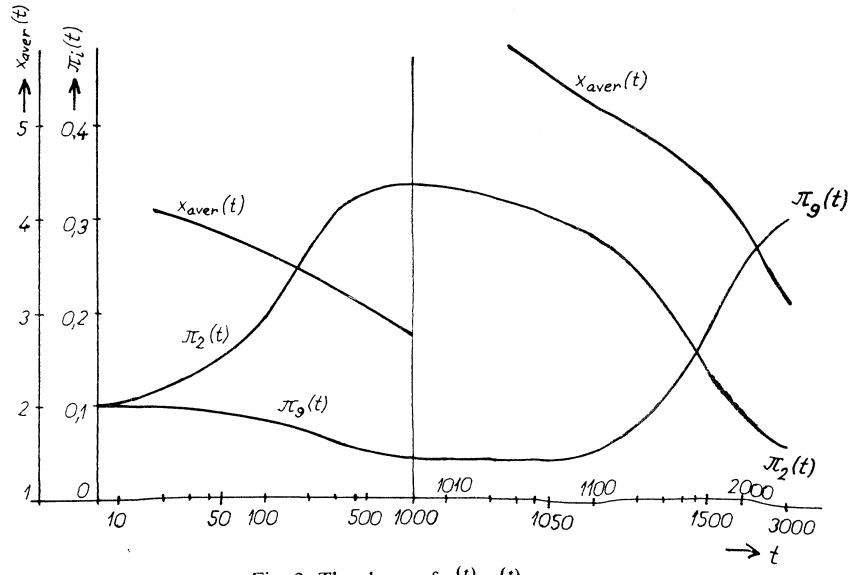

Fig. 2. The shape of $\pi_2^{(t)}$, $\pi_9^{(t)}$, $x_{aver}(t)$.

An integer n is introduced in formula (4). This integer gives the power of the measured value of the controlled variable x. It is obvious that the bigger is n, the more curved shapes of u_i's are obtained, the more conspicuous extremes of u_i's arise and, thus, the more easily SA can recognize the interval with the maximum value of u_i. The dependence of the optimal n on the curvature of the shape of $S(y)$ has been found experimentally [7].

The above mentioned model can be very easily generalized to the N-dimensional case, $N > 1$. By successive counting of the N-dimensional intervals $\mathscr{Y}_i$ we can transform this problem to the one-dimensional case and use the above described algorithm. For details see [6].

3. CONCLUSIONS

The stochastic automaton does not examine the entire shape of the characteristic of the controlled system and does not store it in its memory; it only finds out – according to randomly measured values of the controlled variable – the interval with the minimum mean value of the controlled variable x.

There is a disadvantage of the described model: a rather long time is needed for SA to find the interval with the minimum x. On the other hand a great advantage is that SA does not require a "teacher" and works as an "on-line" regulator so that it gains some experience during its operation and directly controls the given system according to the acquired experience on the shape of the static characteristic of the controlled system.

The extremal regulator stores only the estimates of $\mathsf{E}u_i$, $i = 1, \ldots, l$ inside its inner memory so that its memory has only l cells.

From the formula (8) we can see that the extremal regulator "forgets" the values of the controlled variable measured a long time ago. The reason is that the coefficient α in (8) maintains the same value during the whole operation. But the value of α must not be too small, otherwise undesirable transition would arise in SA.

Other models of SA as learning regulators are in [8], [9], [10], [11], [12], [13]. These models demonstrate that the use of the stochastic approach to learning can discover new methods and insights in this area.

Appendix 1. Stochastic automaton

Definition. *A stochastic automaton of the Moore type* is a 6-tuple

$$A = [\mathscr{Q}, \mathscr{X}, \mathscr{Y}, \{\mathbf{a}(x)\}_{x \in \mathscr{X}}, \mu, \boldsymbol{\pi}] \tag{12}$$

where $\mathscr{Q} = \{q_1, \ldots, q_n\}$ is the set of states,
$\mathscr{X}$ is the input alphabet,
$\mathscr{Y}$ is the output alphabet,
$\mathbf{a}(x) = [a_{ij}(x)]_{n,n}$ is a stochastic matrix, called transition matrix,

$\mu : \mathcal{Q} \to \mathcal{Y}$ is the so-called mark function (deterministic or random function),
π is the vector of initial probabilistic distribution of the states,

with the interpretation

$$a_{ij}(x) = \mathsf{P}(q(k+1) = q_j \mid q(k) = q_i,\ x(k) = x) \tag{13}$$
$$y(k) = \mu(q(k+1)),\ k = 1, 2, \ldots .$$

Here $q(k)$ is the state of the automaton in the step k, $x(k)$ ($y(k)$) is the input (output) symbol of the automaton in the step k, $\mathsf{P}(\ldots \mid \ldots)$ means the conditional probability.

SA (stochastic automaton) works as follows: according to the initial distribution π the initial state $q(1)$ of the automaton will be, say, q_i. Let the first input symbol be $x(1) = x_1$. According to the distribution vector

$$\pi(2) = \mathbf{a}(x_1)\,\pi \tag{14}$$

SA transfers randomly to a state $q(2)$, say q_j, and yields the output symbol $y(1) = \mu(q_j)$, etc.

The theory of stochastic automata allows us to describe the behaviour of an automaton not only by the transition matrices $\mathbf{a}(x)$ but also by a series of vectors $\pi(k)$ of the probabilistic distribution of states:

$$\pi(k) = [\pi_1(k), \ldots, \pi_n(k)],\quad k = 1, 2, \ldots \tag{15}$$

where $\pi_i(k)$ is the probability that SA will be within the state q_i at step k, $i = 1, \ldots, n$.

This description is used chiefly when SA has a variable structure, i.e. its transition matrices depend on a current step k:

$$\{\mathbf{a}(k, x)\}_{x \in \mathcal{X}},\quad k = 1, 2, \ldots .$$

Appendix 2. Stochastic approximations

Theorem 1 presented above is derived from the following two theorems: the former is the well-known Dvoretzky's Theorem on stochastic approximations, the latter is an original one.

Theorem 2. (*Dvoretzky*, [5]). *Let $Z_0, X_1, X_2, \ldots$ be random variables, let*

$$Z_{n+1} = T_{n+1}(Z_0, \ldots, Z_n) + X_{n+1},\quad n = 0, 1, \ldots \tag{16}$$

let θ be a real number and, furthermore, let

a) *T_{n+1}, $n = 0, 1, \ldots$ be a measurable function satisfying*

$$(\forall r_0, \ldots, r_n)\,|T_{n+1}(r_0, \ldots, r_n) - \theta| \leqq c_{n+1}|r_n - \theta| \tag{17}$$

where $\{c_n\}_{n=1}^{\infty}$ is a series of positive numbers so that

$$\prod_{n=1}^{\infty} c_n = 0 \tag{18}$$

b) $\mathsf{E}Z_0^2 < \infty$,

c) $\mathsf{E}(X_{n+1} \mid Z_0, \ldots, Z_n) = 0$ *with probability* 1, $n = 0, 1, \ldots$,

d) $\sum_{n=1}^{\infty} \mathsf{E}X_n^2 < \infty$.

Then

$$\mathsf{P}(\lim_{n\to\infty} Z_n = \theta) = 1\,, \quad \lim_{n\to\infty} \mathsf{E}(Z_n - \theta)^2 = 0\,. \tag{19}$$

Theorem 3. *Let* $Z_0, X_1, X_2, \ldots$ *be independent random variables with* $\mathsf{E}X_n = \mu$, $\mathsf{D}X_n \leqq D$, $n = 1, 2, \ldots$, $\mathsf{E}Z_0 = \mu$, $\mathsf{D}Z_0 \leqq D$. *Let*

$$Z_{n+1} = \alpha Z_n + (1 - \alpha) X_{n+1}\,, \quad n = 0, 1, \ldots \tag{20}$$

where $0 < \alpha < 1$. *Then*

1) $\mathsf{E}Z_n = \mu\,, \quad n = 0, 1, \ldots\,,$

2) $(\forall \varepsilon > 0) \lim_{n\to\infty} \mathsf{P}(|Z_n - \mu| > \varepsilon) \leqq \frac{D}{\varepsilon^2} \frac{1 - \alpha}{1 + \alpha}\,,$

3) $\mathsf{P}\left(\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} Z_i = \mu\right) = 1\,, \quad \lim_{n\to\infty} \mathsf{E}\left(\frac{1}{n+1} \sum_{i=0}^{n} Z_i - \mu\right)^2 = 0\,.$

Proof.

1) Formula (20) yields

$$Z_n = \alpha^n Z_0 + (1 - \alpha)\alpha^{n-1} X_1 + \ldots + (1 - \alpha) X_n = \tag{21}$$

$$= \alpha^n Z_0 + (1 - \alpha) \sum_{i=1}^{n} \alpha^{n-i} X_i\,, \quad n = 0, 1, \ldots$$

$$\mathsf{E}Z_n = \mu\left[\alpha^n + (1 - \alpha) \sum_{i=1}^{n} \alpha^{n-i}\right] = \mu\left[\alpha^n + (1 - \alpha) \frac{1 - \alpha^n}{1 - \alpha}\right] = \mu\,.$$

2) According to (21), using the independence of $Z_0, X_1, X_2, \ldots$ we can get ($n = 0, 1, \ldots$)

$$\mathsf{D}Z_n \leqq D\left[\alpha^{2n} + (1 - \alpha)^2 \sum_{i=1}^{n} \alpha^{2(n-i)}\right] = D\left[\alpha^{2n} + (1 - \alpha)^2 \frac{1 - \alpha^{2n}}{1 - \alpha^2}\right] = \tag{22}$$

$$= D \frac{1 - \alpha + 2\alpha^{2n+1}}{1 + \alpha}\,.$$

The Chebyshev inequality yields

$$(\forall \varepsilon > 0)\ \mathsf{P}(|Z_n - \mu| > \varepsilon) < \frac{D}{\varepsilon^2} \frac{1 - \alpha + 2\alpha^{2n+1}}{1 + \alpha}\,.$$

Finally,

$$(\forall \varepsilon > 0) \lim_{n\to\infty} \mathsf{P}(|Z_n - \mu| > \varepsilon) \leqq \frac{D}{\varepsilon^2} \frac{1 - \alpha}{1 + \alpha}\,.$$

3) Let us assign

$$W_n = \frac{1}{n+1} \sum_{i=0}^{n} Z_i , \quad n = 0, 1, \ldots .$$

The recurrent formulas for W_n are

$$W_0 = Z_0 ,$$

$$W_{n+1} = \left(1 - \frac{1}{n+2}\right) W_n + \frac{1}{n+2} Z_{n+1} , \quad n = 0, 1, \ldots .$$

By subtracting μ from both sides of this equation and by assigning

$$\overline{W}_n = W_n - \mu , \quad \overline{Z}_n = Z_n - \mu , \quad n = 0, 1, \ldots$$

we get

$$\overline{W}_{n+1} = \left(1 - \frac{1}{n+2}\right) \overline{W}_n + \frac{1}{n+2} \overline{Z}_{n+1} , \quad n = 0, 1, \ldots .$$

Now let us prove that the assumptions of Dvoretzky's Theorem are fulfilled.

a) Let

$$T_{n+1}(r_0, \ldots, r_n) = \left(1 - \frac{1}{n+2}\right) r_n , \quad n = 0, 1, \ldots .$$

Thus

$$\left|T_{n+1}(r_0, \ldots, r_n)\right| = \left(1 - \frac{1}{n+2}\right) |r_n| , \quad n = 0, 1, \ldots .$$

But

$$\prod_{n=1}^{\infty} \left(1 - \frac{1}{n+1}\right) = 0 .$$

The formula (18) is thus valid.

b) $\mathsf{E}\overline{W}_0^2 = \mathsf{D}Z_0 \leqq D < \infty$,

c) $\mathsf{E}(\overline{Z}_{n+1} \mid \overline{W}_0, \ldots, \overline{W}_n) = 0 , \quad n = 0, 1, \ldots ,$

d) $\sum_{n=1}^{\infty} \mathsf{E}\left(\frac{1}{n+1} \overline{Z}_n\right)^2 \leqq D \sum_{n=1}^{\infty} \frac{1}{(n+1)^2} < \infty$

because according to (22)

$$\mathsf{E}\overline{Z}_n^2 = \mathsf{D}Z_n \leqq D \frac{1 - \alpha + 2\alpha^{2n+1}}{1 + \alpha} \leqq D , \quad n = 1, 2, \ldots .$$

The assumptions of Dvorezky's Theorem are fulfilled for $\theta = 0$ and thus we obtain

$$\mathsf{P}(\lim_{n \to \infty} W_n = \mu) = 1 , \quad \lim_{n \to \infty} \mathsf{E}(W_n - \mu)^2 = 0$$

QED.

References

[1] *A. Paz:* Introduction to probabilistic automata. Academic Press, New York and London 1971.

[2] *М. Л. Цетлин:* О поведении конечных автоматов в случайных средах. Автоматика и телемеханика 22 (1961), 1345—1354.

[3] *В. И. Варшавский, И. П. Воронцова:* О поведении стохастических автоматов с переменной структурой. Автоматика и телемеханика 24 (1963), 353—360.

[4] *K. S. Fu, T. J. Li:* Formulation of learning automata and automata games. Information Sciences 1 (1969), 237—256.

[5] *A. Dvoretzky:* On stochastic approximation. Proc. 3rd Berkeley Symp. Math. Statist. and Probability, vol. 1, 39—55, Univ. of California Press, Berkeley, Cal., 1956.

[6] *I. Brůha:* Comparing the theory of deterministic and probabilistic automata for modelling adaptive learning systems (Czech). Ph. D. thesis, FEL ČVUT, 1973.

[7] *P. Benedikt:* Modelling learning systems by means of probabilistic automata (Czech). Master Thesis, FEL ČVUT, 1974.

[8] *K. S. Fu:* Stochastic automata as models of learning systems. Proc. Symp. Comp. Information Sci., Columbus, Ohio, 1966.

[9] *K. S. Fu, Z. J. Nikolic:* On some reinforcement techniques and their relation to the stochastic approximation. IEEE Trans. AC-11 (1966), 756—758.

[10] *K. S. Narendra, M. A. L. Thathachar:* Learnig automata — a survey. IEEE Trans. SMC-4 (1974), 323—334.

[11] *Y. Sawaragi, N. Baba:* Two ε-optimal nonlinear reinforcement schemes for stochastic automata. IEEE Trans. SMC-4 (1974), 126—131.

[12] *R. Viswanathan, K. S. Narendra:* Games of stochastic automata. IEEE Trans. SMC-4 (1974), 131—135.

[13] *Z. Kotek, I. Brůha, V. Chalupa, J. Jelínek:* Adaptive and learning systems (Czech). SNTL Praha, 1980.

Souhrn

UČÍCÍ SE EXTREMÁLNÍ REGULÁTOR MODELOVANÝ POMOCÍ PRAVDĚPODOBNOSTNÍHO AUTOMATU A TEORIE STOCHASTICKÝCH APROXIMACÍ

IVAN BRŮHA

Vlastnosti učících se systémů lze studovat z mnoha přístupů, využívaje různých oblastí matematiky. Výsledky však obvykle byly příliš komplikované či nesouhlasily s výsledky praktických pokusů.

Článek uvádí modelování učících se systémů pomocí stochastických automatů. Podrobně je vyvětlen jeden model učícího se extremálního regulátoru. Důkaz konvergence je založen na Dvoretzkyho větě o stochastických aproximacích. Ukazuje se, že stochastické automaty s teorií stochastických aproximací jsou vskutku vhodným nástrojem pro studium učících se systémů.

Author's address: RNDr. Ing. *Ivan Brůha,* CSc., Elektrotechnická fakulta ČVUT, Suchbátarova 2, 166 27 Praha 6.