ON THE ADAPTIVE IDENTIFICATION OF SYSTEMS WITH TIME-VARYING PARAMETERS

By

I. VAJK and L. KEVICZKY

Department of Automation, Technical University, Budapest
(Received January 11, 1975)

Presented by Prof. Dr. F. CSÁKI

Introduction

The adaptive identification of systems with time-varying parameters and varying environment is often attributed to the extremization of functional

$$J^*(\mathbf{c}, t) = M_{\mathbf{x}}\{Q(\mathbf{x}, \mathbf{c}, t)\} \tag{1}$$

where the distribution $p_{\mathbf{x}}(\mathbf{x}, t)$ of function $Q(\mathbf{x}, \mathbf{c}, t)$ is not known. Here $\mathbf{x}$ changes according to a random process, $\mathbf{c}$ is the vector of unknown parameters and t means the time. Unfortunately the functional (1) cannot be used directly in most of identification procedures since it is not completely determined. That is why in many cases it is empirically estimated.

Two, most often used approximations:

1. $$J^*(\mathbf{c}(t), t) = \int_0^t w(t, \tau)\, Q\big(\mathbf{x}(\tau), \mathbf{c}(t), \tau\big)\, d\tau\,, \tag{2}$$

where the parameter changes are taken into consideration by the weighting function $w(t, \tau)$ [4], [5].

2. $$J^*(\mathbf{c}(t), t) = \int_0^t Q\big(\mathbf{x}(\tau), \mathbf{c}(\mathbf{b}, \tau), \tau\big)\, d\tau\,, \tag{3}$$

where $\mathbf{c}(\mathbf{b}, \tau)$ is known except the case $\mathbf{b} = \text{const}$ [1], [2], [3], [6]. This paper is concerned with the determination of weighting function of functional (2).

The necessity of weighting

The on-line identification methods based on weighting permit to follow the changes of the system and its environment by gradually changing the model parameters. The adaptation is concomitant to forgetting the previous data, since their information content is less than that of the actually measurements.

To forget may be necessary in the following cases:

1. If there is no difference between the structure of the system and the model, the parameters are constant in time, and the observations are only disturbed by random noise, then the weighting is used in order to improve the stochastic convergence of the estimation. It is known from the theory of classical stochastic approximation that in this case weighting series providing for stochastic convergence should be applied.

2. If it is known a *priori* that the points in a given range are disturbed by a greater error then it is reasonable to assign them a lower weight. Here weighting means the unification of noise. Here also an a *priori* known weighting matrix may be applied.

3. In case of time-varying parameters there is a moving target parameter vector which the estimated one has to be converged to, i.e. the stochastic convergence becomes meaningless for infinite time. This really means a simple servo problem in its general sense. Forgetting means the transport of data through filter causing lag and damping in the parameter adaptation. The presence of a noise is against the fast adaptation, since in this case the noise would also be followed. The stochastic convergence must be provided dinamically. Estimation of the trend of parameter change and of the correlation time of noise can be used to determine the speed of forgetting. Necessity of forgetting the previous data is seen by the loss of approximation.

4. The difference of the system and model structure (e.g. in nonlinear systems the changes of workpoint) may impose to forget the data deriving from the previous environment. Rather than from the loss of approximation alone, the necessity of forgetting is also seen by the change of statistical characteristics (expected values, standard deviation) of the input signals. This feed forward allows faster adaptation.

Thus, the difference signal of the forgetting mechanism as an adaptive system can be formed according to the above considerations.

Weighting strategies

In discrete case the functional (2) is of the form:

$$J^*(\mathbf{c}[n], n) = \sum_{k=0}^{n} w(n, k)\, Q(\mathbf{x}[k], \mathbf{c}[n], k)\,, \tag{4}$$

where $w(n, k)$ is a suitably chosen weighting function. In stationary case weighting is made as:

$$w(n, k) = \frac{1}{n}. \tag{5}$$

In identifying time-varying parameters $w(n, k)$ is often chosen as

$$w(n, k) = \prod_{i=k+1}^{n} d[i] \tag{6}$$

and

$$w(n, n) = 1 ,$$

where $d[i]$ is forgetting factor at a time i. Such a choice of the weighting function leads to functional

$$J(\mathbf{c}[n], n) = nJ^*(\mathbf{c}[n], n) \tag{7}$$

instead of (4), but the extremums of these two functionals are identical within the parameter range. Using the forgetting factor, functional (4) can be written as:

$$J(\mathbf{c}[n], n) = d[n]\, J(\mathbf{c}[n], n - 1) + Q(\mathbf{x}[n], \mathbf{c}[n], n) . \tag{8}$$

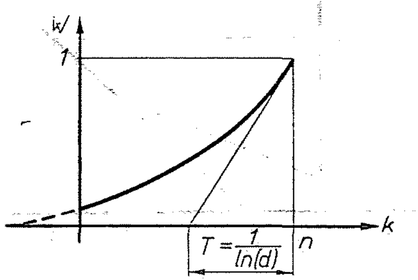


Fig. 1

The exponential, linear, combined block by block weighting can be discussed as special cases of the above mentioned general one.

1. In exponential weighting $d[i] = d =$ const. In the functional.

$$J(\mathbf{c}[n], n) = \sum_{k=1}^{n} d^{n-k} Q(\mathbf{x}[k], \mathbf{c}[n], k) \tag{9}$$

the weighting function is a geometric series which corresponds to an exponential function slope (Fig. 1).

2. In linear weighting the absolute weight of the n-th and $n - 1$-th observations is

$$w(n, n) = 1 \quad \text{and} \quad w(n - 1, n - 1) = \frac{n + m - 1}{n + m} , \tag{10}$$

to be read directly from Fig. 2. Hence the forgetting factor $d[n]$:

$$d[n] = \frac{w(n-1, n-1)}{w(n, n)} = \frac{n+m-1}{n+m}. \tag{11}$$

3. The combined linear weighting is obtained by the recurrent change of parameter m of linear weighting (Fig. 3).

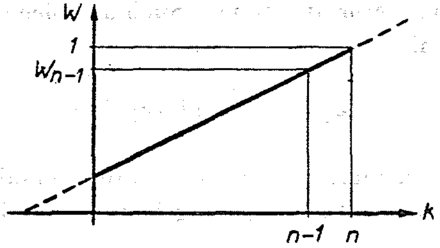


Fig. 2

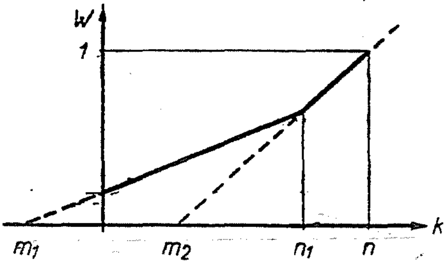


Fig. 3

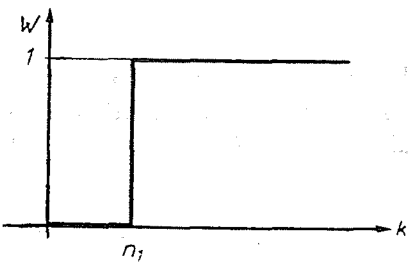


Fig. 4

4. The block by block weighting may be considered as a marginal case of the combined linear one; up to n_1 $d = 0$, else $d = 1$, i.e. using the interval without weighting (Fig. 4).

The most often applied strategy is the exponential weighting, since its algorithm is a very simple and efficient one.

Exponential weighting

Certain estimation problems — assuming multiple input single output systems, linear in parameters — can be reduced to the mathematical model

$$y[n] = \mathbf{f}^T(\mathbf{x}[n])\,\mathbf{c}[n] \tag{12}$$

where $\mathbf{x}$ is the input vector, y is the output and $\mathbf{c}$ is the time dependent parameter vector to be identified. The input and output signals of the system can be measured with perturbances ξ and η, respectively. Fig. 5 shows the identification model.

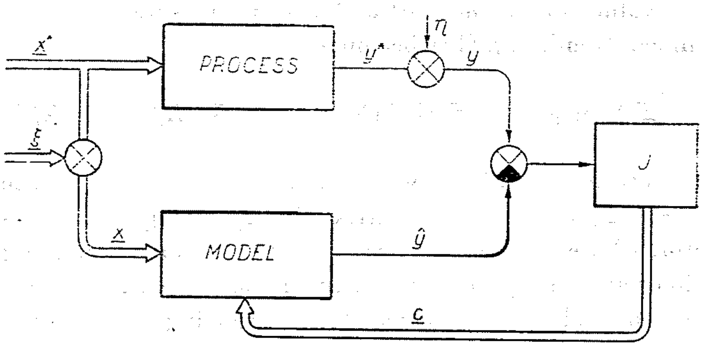


Fig. 5

In order to simplify the notations, a linear model is used and the subscripts of vectors and matrices refer to time, e.g. $\mathbf{f}(\mathbf{x}[n]) = \mathbf{x}_n$.

The functional (4) with exponential forgetting and loss function $Q_n = (y - \hat{y})^2$ is used for identification. In matrix from

$$J(\mathbf{c}_n, n) = (\mathbf{Y}_n - \mathbf{X}_n\mathbf{c}_n)^T\, \mathbf{W}_n(\mathbf{Y}_n - \mathbf{X}_n\mathbf{c}_n)\,, \tag{13}$$

where $\mathbf{Y}_n$ is an $(n \times 1)$ column vector, its components are $y[j]$, $\mathbf{X}_n$ is an $(n \times m)$ matrix with elements $x_{jk} = x_k[j]$, $\mathbf{W}_n$ is an $(n \times n)$ diagonal matrix with elements $w_{jj} = d^{n-j}$, where $0 < d \leq 1$. The unknown vector $\mathbf{c}_n$ is obtained from the extremum of this functional:

$$\mathbf{c}_n = (\mathbf{X}_n^T\mathbf{W}_n\mathbf{X}_n)^{-1}\,\mathbf{X}_n^T\mathbf{W}_n\mathbf{Y}_n\,. \tag{14}$$

The parameter vector can be evaluated recursively using the well-known identities of matrix partition

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mathbf{R}_{n+1}\mathbf{x}_{n+1}(y_{n+1} - \mathbf{x}_{n+1}^T\mathbf{c}_n)\,, \tag{15}$$

where the convergence matrix, optimal in quadratic sense is:

$$\mathbf{R}_{n+1} = \frac{1}{d}\left(\mathbf{R}_n - \frac{\mathbf{R}_n\mathbf{x}_{n+1}\mathbf{x}_{n-1}^T\mathbf{R}_n}{d + \mathbf{x}_{n+1}^T\mathbf{R}_n\mathbf{x}_{n+1}}\right). \tag{16}$$

In regression analysis the determination of parameter d may be difficult.

In several systems the parameters change slowly. Hence the trend of coefficients in a given interval can be considered as linear:

$$\mathbf{c}_n = \boldsymbol{\alpha} + n\boldsymbol{\beta} \tag{17}$$

where $\boldsymbol{\alpha}$ is the value of $\mathbf{c}_n$ at the start and $\boldsymbol{\beta}$ is the measure of parameter change. In case of linear trend, Eq. (14) becomes:

$$\mathbf{c}_n = \left[(\mathbf{X}_n^* + \Xi_n)^T \mathbf{W}_n(\mathbf{X}_n^* + \Xi_n)\right]^{-1}(\mathbf{X}_n^* + \Xi_n)^T \mathbf{W}_n(\mathbf{X}_n^*\boldsymbol{\alpha} + \mathbf{A}_n\mathbf{X}_n^*\boldsymbol{\beta} + \boldsymbol{\eta}_n) \tag{18}$$

according to notations in Fig. 5, where $\mathbf{A}_n$ is an $(n \times n)$ diagonal matrix with elements $a_{jj} = j$, Ξ_n is an $(n \times m)$ matrix with $\xi_{jk} = \xi_k[j]$ and $\boldsymbol{\eta}_n$ is an $(n \times 1)$ column vector, whose components are $\eta[j]$. It is obvious that the goodness of the estimation particularly depends on the choice of parameter d. In order to determine the optimal d the influence of the forgetting factor on the statistical features of the estimation should be investigated.

Further on let us consider the following conditions: the output noise has zero mean $(M\{\eta[j]\} = 0)$, finite variance $\left(M\{\eta^2[j]\} = \sigma_\eta^2 < \infty\right)$ and is uncorrelated $\left(M\{\eta[j]\,\eta[k]\} = 0\right)$. The same is true for the input noise, i.e. $M\{\xi_p[j]\} = 0$, $M\{\xi_p^2[j]\} = \sigma_{\xi p}^2 < \infty$, $M\{\xi_\xi^2[j]\,\xi_p[k]\} = 0$. In addition, the input signals are assumed to be independent with zero mean and finite variance.

Under these conditions the expected value of the parameters can be described as:

$$M\{c_p[n]\} = \frac{\sigma_{x_p}^2}{\sigma_{x_p}^2 + \sigma_{\xi_p}^2}\left(\alpha_p + \frac{\sum\limits_{i=1}^{n} j d^{n-j}}{\sum\limits_{j=1}^{n} d^{n-j}}\,\beta_p\right) \tag{19}$$

where σ_{x_p} is the variance of the p-th input variable.

Eq. (19) shows that in general case the estimation will be biased. The bias depends on the input noise and the trend of parameters. The decrease of variance of input noise reduces the estimation error. The error caused by the trend of parameter depends in particular on the rate of parameter change and on the forgetting factor. If there is no input noise and $d = 1$, then the bias of parameters is: $-\beta_p(n - 1)/2$. If the forgetting factor tends to zero, then the expected parameter value converges to the true value.

But the expectation is not the only feature of the estimation. Its quality may be suitably described by the minimization of the trace of the covariance matrix:

$$V(\mathbf{c}_n) = M\{(\mathbf{c}_n - \mathbf{c}_n^*)(\mathbf{c}_n - \mathbf{c}_n^*)^T\}. \tag{20}$$

Unfortunately its simplification for practical use comes up against difficulties even in linear case. This is why computer simulation is applied to determine the optimum value of the weighting factor *d*.

Results of the simulation investigations

The on-line least-squares algorithm combined with exponential weighting, discussed in the previous section, has been investigated for a complete second-order form. The algorithm was programmed for digital computer and examined for various parameter changes. The program realized the relationships (15) and (16). A few examples will be presented to illustrate the result. The initial values were $\mathbf{R}_0 = 1000\,\mathbf{I}$ and $\mathbf{c}_0 = \mathbf{0}$.

The following sums of square errors served as measure for the goodness of identification:

— for the parameters:

$$S_{ci} = \frac{1}{n} \sum_{k=1}^{n} (c_i[k] - c_i[k])^2 \tag{21}$$

— for the goodness of estimation of y in the i-th period:

$$S_{yi} = \frac{1}{n} \sum_{k=1}^{n} (y_k - \mathbf{c}_k^T \mathbf{f}_k)^2 \tag{22}$$

— and for the average deviation:

$$S_y = \frac{1}{p} \sum_{i=1}^{p} S_{yi} \tag{23}$$

n being the number of steps in one period and p the number of periods.

In the investigations presented here the number of iterations was 400. In linear case the length of a running up or down took 100 steps. In the figures and tables the type of parameter change, their minimum and maximum values separated by ÷ are also indicated.

The simulations show the estimation to depend on the rise of parameter change, the noise level, the variance of the input vector, the number of input variables and observations.

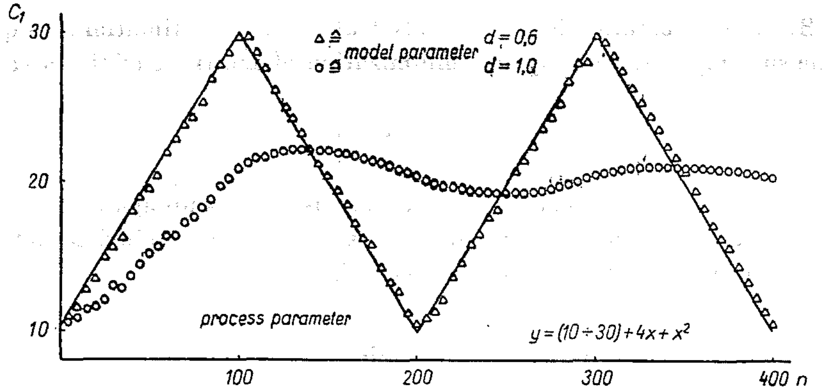


Fig. 6

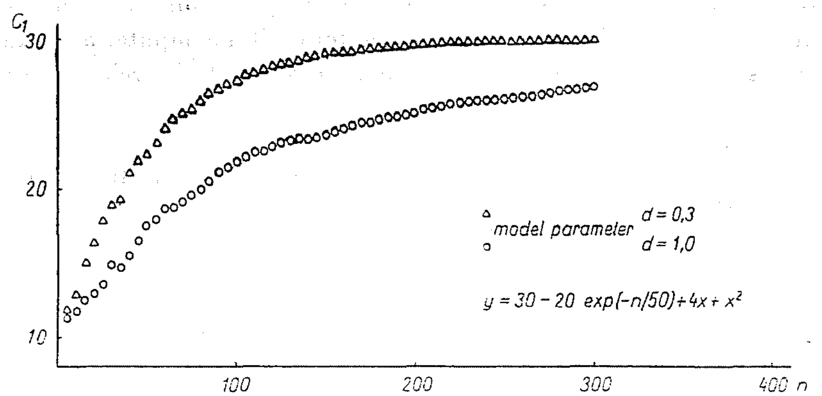


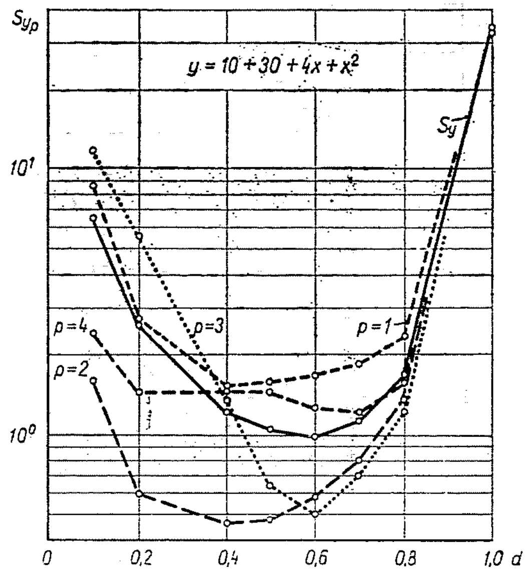
Fig. 7

Figs 6 and 7 show the results of parameter adaptation performed with exponential weighting and without weighting. Tests showed the quality function to have an extremum as a function of forgetting factor (Figs 8 and 9). The convergence of the estimation depends on the size of parameter change. Increasing the size, for linear and sine parameter change Figs 10 and 11 show the value of optimum forgetting factor to decrease. In Table I the values S_y

Table I

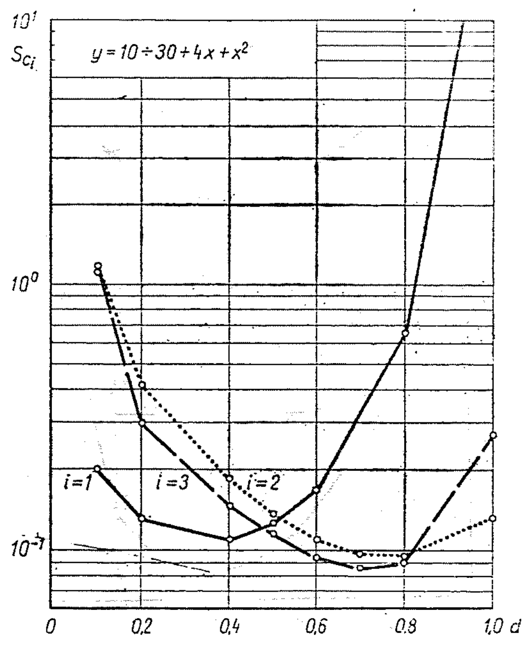
T	S_y	S_{c_0}
100	64.2	2.87
200	2.47	0.200
400	0.442	0.0527
800	0.238	0.00484

$$y = 20-10 \cos (0.0314t) + 4x + x^2$$



S_{yp}
$y = 10 \div 30 + 4x + x^2$
S_y
10^1
$p=4$
$p=3$
$p=1$
$p=2$
10^0
0
0,2
0,4
0,6
0,8
1,0 d

Fig. 8



10^1
S_{ci}
$y = 10 \div 30 + 4x + x^2$
10^0
i=1
i=3
i=2
10^{-1}
0
0,2
0,4
0,6
0,8
1,0 d

Fig. 9

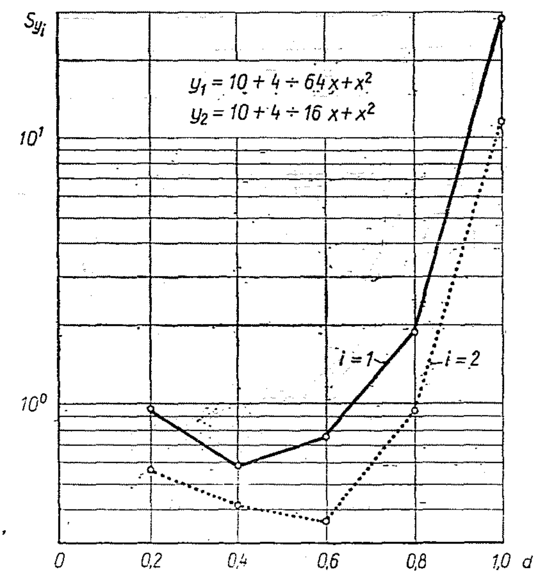

S_{y_i}
$y_1 = 10 + 4 \div 64\,x + x^2$
$y_2 = 10 + 4 \div 16\,x + x^2$
10^1
10^0
i = 1
i = 2
0
0,2
0,4
0,6
0,8
1,0
d

Fig. 10

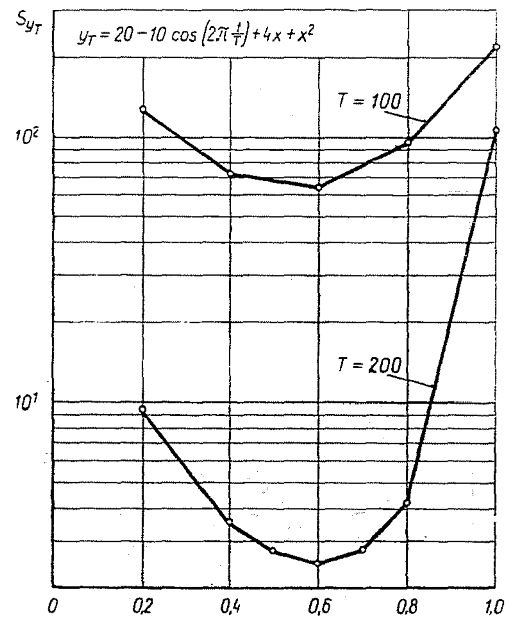

S_{y_T}
$y_T = 20 - 10 \cos\left(2\pi \frac{t}{T}\right) + 4x + x^2$
T = 100
T = 200
10^2
10^1
0
0,2
0,4
0,6
0,8
1,0

Fig. 11

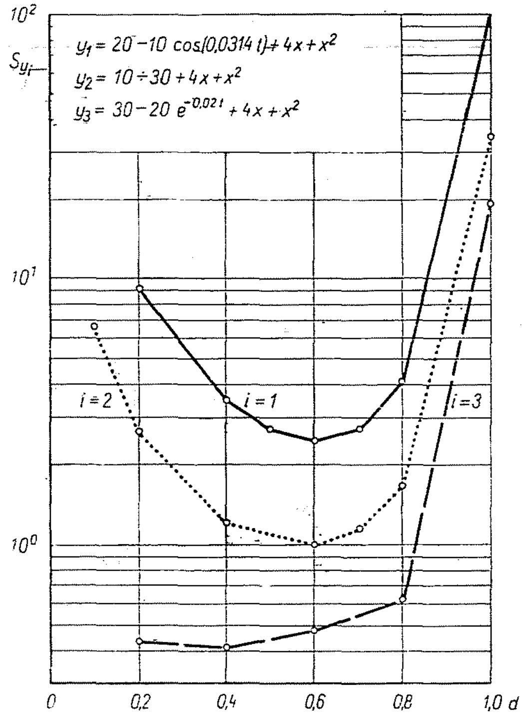


Fig. 12

are presented as a function of period time for sine-varying parameters and $d = 0.6$. Fig. 12 shows the function $S_y(d)$ for parameter changes of various types. Increasing the output noise, the optimal forgetting factor increases (Table II). Goodness of the estimation also depends on the variance of the input vector (Table III). Increasing the number of parameters, the optimal forgetting factor increases. These facts are easy to explain because the decrease of forgetting factor means to "reduce" the number of data used in the estimation. Thus, for a forgetting factor $d = 0.2$ the weight of the 5-th observation is $0.2^4 = 0.0016$.

Table II

σ_y \ d	0.4	0.6	0.8
0	0.412	0.478	0.924
1	1.012	8.129	18.51
10	551.3	155.9	4968

$$y = 10 + (4 \div 64)x + x^2$$

Table III

σ_x \ d	0.4	0.6	0.8
1	0.412	0.478	0.924
10	569	749	1861

$$y = 10 + (4 \div 64)x + x^2$$

The increase of dynamic sum of squares indicates the parameter change. This fact is also of use for determining the optimal forgetting strategy where the increase of the functional without weighting can be used as a difference signal.

Summary

The identification of systems with time-varying parameters is often attributed to the extremization of functional

$$J(\mathbf{c}(t), t) = \int_0^t w(t, \tau)\, Q(\mathbf{x}(\tau), \mathbf{c}(t), \tau)\, d\tau \, ,$$

where the weighting function $w(t, \tau)$ taken into account the parameter changes. This paper deals with the choice of weighting function $w(t, \tau)$, and with the statistical investigation of the estimation and shows the efficiency of the exponential weighting by computer simulation.

References

1. Цыпкин Я. З.: Алгоритмы динамической адаптации. Автоматика и Телемеханика, 1972, № 10.
2. Цыпкин Я. З.: Принципы построения адаптативных и обучающисхя систем. AKI közlemények 1972. 3.
3. Цыпкин Я. З.: Алгоритмы адаптации и обучения в нестационарных условиях. Техническая кибернетика, 1970, № 5.
4. Gyürki, J.: Identification Algorithms in Computer Process Control (in Hungarian). COMPCONTROL-70 Miskolc 1970.
5. Kamen Velec: Choice of the Weighting Coefficient of the Running Regression Analysis with Exponential Weighting Function. IFAC, Hague, 1973.
6. Lee, R.: Optimal Estimation, Identification and Control. MIT Press, Cambridge 1964.

István Vajk
Dr. László Keviczky
} H-1521 Budapest