Engineering Simulation, 1999, Vol. 16, pp. 341—350
Reprints available directly from the Publisher
Photocopying permitted by license only

On a Non–Parametric Algorithm for Smoothing Parameter Control in Adaptive Filtering

S. A. VOROBYEV and E. V. BODYANSKIY

Kharkov State Engineering University of Radioelectronics, Ukraine.

An algorithm for the parameter control of obsolete information discontinuation is proposed on the basis of the filtration problem analysis of random sequences under the conditions of parameter drift. The algorithm is based on using non-parametric methods of mathematical statistics. A criterion for the algorithm development was the frequency of the identification error sign change (renovating sequence).

Key words: adaptive filtering, identification, simulation of stochastic systems.

An identification problem in real time under conditions of parametric drift of the identified model, is associated with checking statistical characteristics of renewing identificator sequence. This is met in many problems of engineering diagnostics, metrology, identification of electron circuits and media. Failures of measuring sensors, sudden shifts, anomalous changes occurring in a measurement channel, changes of statistical characteristics of object noise, failures of the computer and device accuracy decrease result in changing statistical characteristics of renewing identification sequence. This is the consequence of estimation quality deterioration. Under real operating conditions of electric circuits, the problem of operative detection of these changes arises for further correction of estimates calculated by identificator. Consider the simple adaptive algorithm for estimation (filtering) which can control the changes of renewing sequences and efficiently adjust identificator parameters for error correction.

Problem statement and verification. Classic problem of estimation (filtering) is formulated as the problem of seeking unobservable value $\theta(t)$ which is vector one in a general case, partially observable random sequence $(\theta, y) = (\theta(t), y(t))$ (where $t = 0, 1, 2, \ldots$ is the discrete time) where only component $y = (y(t)), t = 0, 1, 2, \ldots,$ is observed, using the values $y_0^t = \{y(s), 0 \le s \le t\}$. The estimate of the unobservable component of the random sequence $\theta(t)$ which is known to be optimal in a mean-square sense is the posterior average value $\hat{\theta}(t) = E\{\theta(t)/y_0^t\}$ where $E\{\cdot\}$ is the symbol of mathematical expectation. The estimate was made with many observations

y_0^t when the conditions $E\{\theta^2(t)\} < \infty$ were met. It is also known that the problem of estimation (filtering) may be described as that of defining the time evolution of the conditional probability density $p(\theta(t)/y_0^t)$ [1, 2].

The designations for partially observable random sequence $(\theta(t), y(t))$ will be interpreted as follows. The vector of model parameters is designated in terms of $\theta(t)$. This model describes the output signal of the control object $y(t)$ and its estimates are designated as $\hat{\theta}(t)$. Let the real random sequence be described by the difference equation in the autoregression form:

$$y(t) = \theta^T \varphi(t) + \omega(t), \tag{1}$$

where $y(t)$ is the observable component of random sequence which is the output object variable (for simplicity, the case of scalar output is considered though the results obtained can be easily generalized for the case of scalar output): $\theta(t) = (a_1(t), a_2(t), ..., a_n(t))^T$ is the vector of unknown process parameters (or the unobservable component of the sequence $(\theta(t), y(t))$); $\varphi(t) = (-y(t-1), -y(t-2), \dots, -y(t-n))^T$ is the vector of previous history; $\omega(t)$ is the random noise of the output signal measurement about which it is known that $E\{\omega(t)\} = 0$, $E\{\omega^2(t)\} = \sigma_\omega^2 < \infty$; n is the model order; T is the transposition symbol.

The vector values of the unknown parameters $\theta(t)$ should be estimated on the basis of the set of observations y_0^t. By (1) the model to be tuned is written as:

$$\hat{y}(t) = \hat{\theta}^T(t-1)\varphi(t), \tag{2}$$

where $\hat{\theta}(t-1) = (\hat{a}_1(t-1), \hat{a}_2(t-1), ..., \hat{a}_n(t-1))^T$ is the vector of tuned estimates of unknown parameters of the random sequence which are evaluated by adaptive identification algorithms. The adaptive identification algorithm [3] used in this case has the form:

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + r^{-1}(t)(y(t) - \hat{\theta}^T(t-1)\varphi(t))\varphi(t), \\ r(t) &= r(t-1) + ||\varphi(t)||^2. \end{aligned} \tag{3}$$

By the solution of filtration problems, the parameters of random sequence are not fixed and are drifting within the filtration interval. Therefore a tracking problem of sequence parameter drift often arises. But the algorithm (3) associated with a group of algorithms for stochastic approximation has the low convergence velocity and cannot be applied when there is drift of the model parameters (1). Under drift conditions the estimation algorithm is used to realize a compromise between filtering and tracking properties. To do this, the procedures with finite memory volume [4] are involved since the increase of the algorithm memory volume results in improving its filtering properties. This is necessary if the parameters $\theta(t)$ remain constant within sufficiently long time intervals. The decrease of the algorithm memory volume improves the tracking properties which is necessary for tracking parameter $\theta(t)$ changes in the case of their drift. Taking into account that it is difficult to choose beforehand a necessary (optimal) constant value of the memory volume for nonstationary random sequence, the procedures of algorithm memory control are forced to be additionally introduced.

The algorithms with controlled memory volume involve, for example, the exponential weighted recursive method of least squares (EWRMLS). But when the synthesis of these algorithms is made, the identification error of a nonstationary object in the case of structural equivalence of object and model consists of two parts: the component caused by the presence of interference in the phase variable which value is inversely proportional to the memory volume, and the component caused by object parameter drift which is in direct relation to the algorithm memory volume.

The available methods of memory control are constructed as a rule according to one principle. The current value of memory volume is defined according to the current value of the criterion to be used which specifies the quality of the filtering process. The direction of the memory volume change is determined by the calculated value of additional criterion by means of its check as to the excess of accessible limits.

An adaptive adjustment procedure is proposed [5] for the algorithm memory volume. It is based on the hypothesis of the directed object parameter drift which is close to the linear one. The adjustment of the algorithm memory volume is performed depending on the value of additional criterion which specifies the hodograph twist of the estimate vector. But this method of memory control is based on the implicit linear drift parametrization. With the essentially nonlinear rule of parameter change, the hodograph twist of the estimate vector may be large due to noise effect rather than because of the object parameter drift. Therefore errors are possible when the directions of the memory volume change are being determined.

An approach to the memory volume control described in [4, 6] is based on studying the statistics which specify errors of the object output prediction. For taking decisions on the expediency of the memory volume increase, observations on checking window of the ψ length are used. With its help, statistics are formed which specify the error of the object output prediction, its distribution being compared with χ^2 distribution, and its properties being used for checking the hypothesis of the absence of the prediction shift or its presence. However, for this approach an assumption of stepwise parameter change is characteristic.

As to EWRMLS, the algorithms of memory control [7—10] are used for protecting identification contour from the explosion of the covariation matrix parameters and solve indequately the problem of drift tracking. Hence the need to detect general drift regularities is a disadvantage of known methods. It gives rise to errors in defining the direction of memory volume change, and to the inconvenience of their numerical implementation. So, it is difficult to realize the comparison of calculated statistics distribution with χ^2 distribution in computational procedure. It gives rise to additional difficulies in the solution of the main problem of parameter estimation. Therefore a procedure of algorithm memory volume adjustment is needed which does not depend on a priori assumptions on the law of the object parameter drift and is simple from the computational standpoint. This algorithm can be constructed using the ideas of [4, 6] and simple non-parametrical methods of mathematical statistics.

Estimation algorithm. Convergence conditions. Consider a stochastic adaptive algorithm for parametric estimation with adjustable memory volume:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{a}{r(t)}(y(t) - \hat{\theta}^T(t-1)\varphi(t))\varphi(t),$$

$$r(t) = \alpha(t) r(t-1) + \|\varphi(t)\|^2,\ 0 \le \alpha(t) \le 1,\ 0 < a < 2, r(0) = 1, \tag{4}$$

where α is a certain constant value.

The difference of the algorithm (4) from (3) is in using the discount parameter $\alpha(t)$. It is involved in (4) on the basis of the fact that $r(t)$ is a single member which takes into account previous history of the algorithm (3). Consequently, the introduction of the parameter $\alpha(t)$ would permit the adjustment of the memory volume of the estimation algorithm by itself. With $\alpha(t) = 0$ the algorithm (4) demonstrably coincides with the optimal one-step algorithm [11]

$$\theta(t) = \hat{\theta}(t-1) + \frac{y(t) - \hat{\theta}^T(t-1)\varphi(t)}{\|\varphi(t)\|^2}\varphi(t) \tag{5}$$

and with $\alpha(t) = 1$ it coincides with the adaptive algorithm (3).

We shall make a convergence analysis of the above algorithm (4). It is written relative to estimation error as

$$\tilde{\theta}(t) = \tilde{\theta}(t-1) + \frac{a}{r(t)}\varphi(t)\vartheta_I(t),$$

where $\tilde{\theta}(t) = \theta(t) - \hat{\theta}(t)$, $\vartheta_I(t) = y(t) - \hat{\theta}^T(t-1)\varphi(t) = y(t) - \hat{y}(t)$ and the Lyapunov function is introduced

$$V(t) = \tilde{\theta}^T(t)\tilde{\theta}(t) = V(t-1) + \frac{2a}{r(t)}\tilde{\theta}^T(t-1)\varphi(t)(\vartheta_I(t) - \omega(t)) +$$

$$+ \frac{2a}{r(t)}\tilde{\theta}^T(t-1)\varphi(t)\omega(t) + \frac{a^2}{r^2(t)}\|\varphi(t)\|^2(\vartheta_I(t) - \omega(t))^2 +$$

$$+ 2\omega(t)(\vartheta_I(t) - \omega(t)) + \omega^2(t). \tag{6}$$

By designating $\tilde{\omega}(t) = \vartheta_1(t) - \omega(t)$, $b(t) = -\varphi^T(t)\tilde{\theta}(t-1)$ and by averaging (6) by $\omega(t)$ we obtain

$$E\{V(t)/y_0^{t-1})\} = V(t-1) - \frac{2a}{r(t)} b(t)\tilde{\omega}(t) + \frac{a^2}{r^2(t)} ||\varphi(t)||^2 \tilde{\omega}^2(t) + $$

$$+ \frac{a^2}{r^2(t)} ||\varphi(t)||^2 \sigma_\omega^2 . \tag{7}$$

The further study of convergence based on considering the sequence (7) as the supermartingnal, completely coincides with the approach in [3]. Here the satisfying of the condition $r^2(t) > r(t)\, r(t-1)$ or $(\alpha(t)\, r(t-1) + ||\varphi(t)||^2)^2 > (\alpha(t)\, r(t-1) + ||\varphi(t)||^2)\, r(t-1)$ at each iteration is an important moment.

Consequently, the finding of the required $\alpha(t)$ is associated with the inequality solution

$$\alpha^2(t) + \alpha(t)\frac{2r(t-1)\,||\varphi(t)||^2}{r^2(t-1)} + \frac{||\varphi(t)||^4 - ||\varphi(t)||^2 r(t-1)}{r^2(t-1)} > 0$$

which may be written as $|\beta(t)| > \sqrt{m(t)}$, where

$$\beta(t) = \alpha(t) + \frac{2\,||\varphi(t)||^2 - r(t-1)}{2r(t-1)}, \tag{8}$$

$$m(t) = \left(\frac{2\,||\varphi(t)||^2 - r^2(t-1)}{2r(t-1)}\right)^2 - \frac{||\varphi(t)||^4 - ||\varphi(t)||^2 r(t-1)}{r^2(t-1)} = \frac{1}{4}.$$

From (8) we obtain

$$\alpha(t) + \frac{2\,||\varphi(t)||^2 - r(t-1)}{2r(t-1)} > \frac{1}{2},$$

and have

$$1 - \frac{||\varphi(t)||^2}{r(t-1)} < \alpha(t) \le 1. \tag{9}$$

From (9) the relationship $(1-\alpha(t))\sum_{j=1}^{t} \alpha^{t-j-1}\,||\varphi(t)||^2 \le ||\varphi(t)||^2$, follows which is the solution of the equation $\alpha(t) = \alpha_0\,\alpha(t-1) + (1-\alpha_0),\ 0 \le \alpha_0 \le 1$.

Thus, the algorithm (4) convergence is provided either by constant increase of the value $||\varphi(t)||^2$ which may be done in an artificial way or by the appropriate discount parameter change which is increased from 0 to 1.

The influence of $\alpha(t)$ on the algorithm convergence velocity is estimated. To do this (6) is written as

$$V(t) = V(t-1) - \frac{2a}{r(t)}\tilde{\theta}^T(t-1)\varphi(t)\vartheta_I(t) + \frac{a^2}{r^2(t)}\vartheta_I(t)\,||\varphi(t)||^2.$$

If the Lyapunov function change at each step is taken as the characteristic of the convergence velocity, the solution of the differential equation

$$\frac{(\partial V(t-1) - V(t))}{\partial r(t)} = \frac{2a\tilde{\theta}^T(t-1)\varphi(t)\vartheta_I(t)}{r^2(t)} + \frac{2a^2\vartheta_I^2(t)\,||\varphi(t)||^2}{r^3(t)} = 0$$

permits the value $r(t)$ to be obtained which provides maximum convergence velocity in the form $r(t) = a\,||\varphi(t)||^2$. But $r(t)$ is determined by the second relationship (4) whence it is seen that the maximum algorithm speed is achieved with $a = 1, \alpha(t) = 0$, i.e. it is provided by the procedure (5).

It is known that one-step algorithms to which the procedure (5) also belongs, operate poorly when there are defects, providing convergence only into a certain domain. These algorithms are optimal algorithms for identification of determined objects as to their speed. In order to estimate the sizes of the domain (7) is written as

$$E\{V(t)/y_0^{t-1})\} = V(t-1) - \frac{2a}{||\varphi(t)||^2}\tilde{\theta}^T(t-1)\,\varphi(t)(\vartheta_I(t) - \omega(t)) + $$

$$+\frac{(\vartheta_I(t) - \omega(t))^2}{||\varphi(t)||^2} + \frac{\sigma_\omega^2}{||\varphi(t)||^2},$$

whence $2\tilde{\theta}^T(t-1)\varphi(t)(\vartheta_I(t) - \omega(t)) > (\vartheta_I(t) - \omega(t))^2 + \sigma_\omega^2$, or $(\vartheta_I(t) - \omega(t))^2 > \sigma_\omega^2$ which coincides with the convergence domain of the algorithm (5).

It is clear that the tracking properties of the algorithm (4) contradict to its filtering properties. Therefore during the estimation process it is appropriate to start with a small value of the parameter $\alpha(t)$ providing high convergence velocity and then to increase it up to the level which provides the compromise between filtering and tracking properties of the algorithm. To do this a new procedure of the parameter $\alpha(t)$ control is introduced.

Tuning of the smoothing parameter $\alpha(t)$. which determines the depth of the algorithm memory is as follows.

A parameter drift of random sequence is represented in the sign of identification error (renovating sequence) $\tilde{y}(t) = y(t) - \hat{y}(t)$. It is not changed if there is drift and is undetermined if there is no drift. The algorithm of the smoothing parameter control of the obsolete information is based on the fact that when there is drift of the random sequence parameters, the conditional density of probability distribution $p\,\theta(t)/y_0^t)$ is changed. Consequently, the definition problem of the algorithm memory depth is rigidly bound with that of distribution identity for two random sequences (where one is the real random sequence and the other is created by the inputs of the tuned model (2)). The problem can be solved using the parametric methods of mathematical statistics [12].

With the peculiarities of the above problem taken into account, the modification of Mann-Whitney criterion is most natural and simple from the computational point of view. Its critical domain may be represented in the following form:

$$\left\{ \sum_{i=t-\psi+1}^{t} \operatorname{sgn}(y(i) - \hat{y}(i)) \geq \delta \right\}$$

where δ is a certain allowable limit; ψ is the width of the checking window;

$$\operatorname{sgn}(y(i) - \hat{y}(i)) = \begin{cases} 0, & y(t) = \hat{y}(i), \\ +1, & y(t) > \hat{y}(i), \\ -1, & y(t) < \hat{y}(i). \end{cases}$$

In this case the procedure of parameter $\alpha(t)$ control begins with assigning the value 0 to it as the high velocity of the algorithm convergence is provided. During the identification process the following situations can arise:

$$\left| \sum_{i=t-\psi+1}^{t} \operatorname{sgn}(y(i) - \hat{y}(i)) \right| \leq \delta$$

which means the predominance of the "noise" component of error over "drifting". Here the solution is taken that the assumed value of a posteriori probability density $p(\theta(t)/y_0^t)$ is correct. Thus it is necessary to improve the filtering properties of the estimation algorithm. To do this, the parameter of the algorithm memory volume should be increased, i.e. $\alpha(t+1) = \alpha(t) + \Delta\alpha$, where $\Delta\alpha$ is the constant number, $\Delta\alpha \geq 0$; $\left|\sum_{i=t-\psi+1}^{t} \operatorname{sgn}(y(i) - \hat{y}(i))\right| > \delta$. This points to rather fast drift and degradation of identification quality when the memory volume is increased. Here the solution is taken that the assumed a posteriori probability density $p(\theta(t)/y_0^t)$ is not correct, and consequently it is necessary to improve the tracking properties of the estimation algorithm. For this aim the parameter of the algorithm memory volume should be decreased, i.e. $\alpha(t+1) = \alpha(t) - \Delta\alpha$.

Thus the operation of the above algorithm starts from small values of the $\alpha(t)$ parameter, which provides high convergence velocity of the parameter estimates to

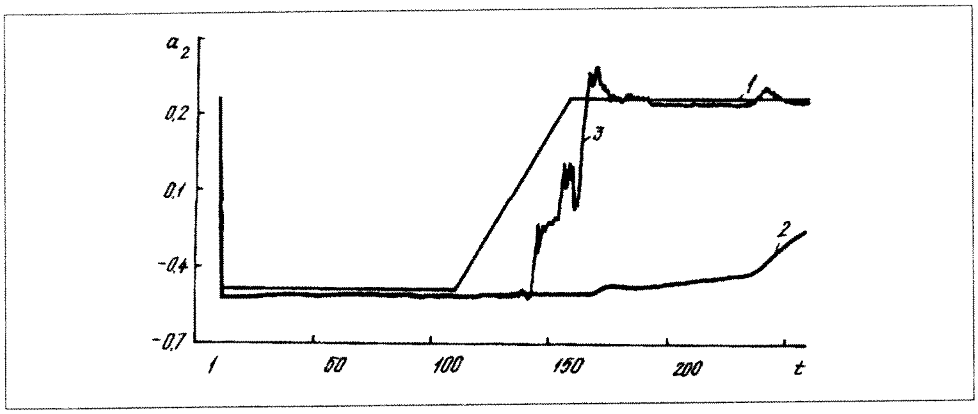

Figure 1. The plots of the parameter a_2 change; 1 — by the model (10), 2 — by the algorithm (3), 3 — by the algorithm (4).

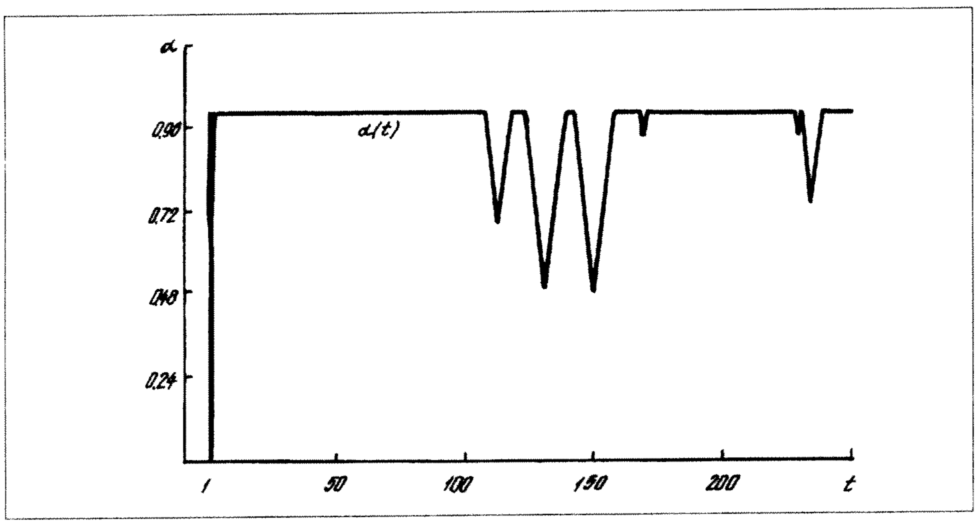

Figure 2. The plot of the smoothing parameter $\alpha(t)$ change.

their values. Then the solution is taken on the increase of the algorithm memory volume as the algorithms with indefinitely increasing memory volume have the best filtering properties. When there is drift of the parameters to be estimated, it becomes necessary to realize compromise between filtering and tracking properties of the estimation algorithm and the solution is taken to decrease the algorithm memory. After drift "compensation" the memory volume again begins to increase.

For successful operation of this algorithm it is natural to assume that the parameter changes occur rather rarely. The solution on the memory volume change by finite access of the identification error signs, width ψ is correct in the case where the previous signs of the identification error describe the sequence distribution densities before the parameter drift appearance. Consequently they give no information on the correct solution. For this same reason it is not necessary to increase the previous history volume if the solution is taken on increasing the parameter of the memory depth control. The estimation quality in this case is evidently worse until the algorithm can take into account the previous history of the necessary duration. It should be noted that the choice of the boundary δ value assumes a certain element of subjectivism.

Example. Let us give a numerical experiment for the estimation of the parameters of random sequence generated by the autoregressive model of the second order. It is given by the following difference equation:

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \omega(t), \tag{10}$$

where a_1, a_2 are the initial values of model parameters; $a_1 = 0.75, a_2 = -0.5; \omega(t)$ is the noise with the following characteristics: mathematical expectation $E\{\omega(t)\} = 0$; mean square deviation $\sigma_\omega = 0.15$. Beginning with the 100-th simulation step, the parameter a_2 is drifting by the law $a_2(t+1) = a_2(t) + h$ where $h = 0.008$. The tuned model corresponding to (10) is written as $\hat{y}(t) = \hat{a}_1(t-1)y(t-1) + \hat{a}_2(t-1y(t-2)$.

The parameters of the tuned model are estimated using the algorithms (3) and (4) with the smoothing parameter defined by the above algorithm. The initial values of the parameter estimates for the tuned model coincide with those of the model (10) parameters.

Fig. 1 presents the plots of the parameter a_2 change by the model (10) (curve 1), by the algorithm (3) (curve 2) and by the algorithm (4) with the parameter $\alpha(t)$ defined by the algorithm proposed (curve 3). Fig. 2 gives the plot of the smoothing parameter $\alpha(t)$ change. The following characteristics of the algorithm for the smoothing parameter control are used in the example: $\psi = 15, \delta = 5, \Delta\alpha = 0.3$.

From Figs. 1 and 2 it is seen that under the conditions of parameter drift, the algorithm proposed provides the convergence to the real parameter value with a small delay while the algorithm (3) does not provide the convergence. This justifies the theoretical result. This algorithm is simple from the computational point of view and is efficient by solving the filtering (estimation) problems under the conditions of parameter drift.

REFERENCES

1. *Filtratsiya i stokhsticheskoe upravlenie v dinamicheskikh sistemakh* (Filtration and stochastic control in dynamic systems)/Ed. K. T. Leondes. — M. :Mir, 1980. — 408 p. (In Russian).
2. Liptser R. Sh., Shirayev A. N. *Statistika sluchainykh protsessov. Nelineinaya filtratsiya i cmezhnye voprocy* (Statistics of random processes. Nonlinear filtering and adjacent problems). — M.: Nauka, 1974. — 696 p. (In Russian).
3. Goodwin G. C., Ramadge P. J., Caines P. E. A globally convergent adaptive predictor//Automatica. — 1981. — 17, No 1. — p. 135—140.
4. Shilman S. V. *Metody adaptivnoi filtratsii sluchainykh protsessov* (Methods of adaptive filtering of random processes)//Dinamicheskie sistemy:adaptatsiya i optimizatsiya. — Gorkiy:Gorkovskiy gos. univ,1985. — p. 22—51. (In Russian).
5. Perelman I. I. *Operativnaya identifikatsiya objektov upravleniya* (Routine identification of control objects). — M.: Energoizdat, 1982. — 272 p. (In Russian).
6. Shilman S. V. *Adaptivnaya filtratsiya vremennykh ryadov* (Adaptive filtering of time series). — N. Novgorog. Izd. Nizhegorodskogo univ., 1995. — 190 p. (In Russian).
7. Fortesque T. R., Kershenbaum L. S., Ydstie B. E. Implementation algorithms with forgetting factor//Automatika. — 1983. — 17, No 6. — p. 831—835.
8. Harrison P. I., Johnston F. R. Discount weighted regression// J. Oper. Res. Soc. — 1984. — 35, No 10. — p. 926—932.
9. Parkum J. E., Poulsen N. K., Holst J. Recursive forgetting algorithm//Int. J. Control. — 1992. — 55, No 1. — p. 109—128.
10. Etxebarria V. Adaptive control with a forgetting factor with multiple samples between parameter adjustment// Int. J. Control. — 1992. — 55, No 5. — p. 1189—1200
11. Kaczmarz S. Approximate solution of systems of linear equations//Int. J. Control. — 1993. — 57, No 5. — p. 1269—1271.
12. Devroy L., Dierfie L. *Neparametricheskoe otsenivanie plotnosti* (Nonparametric density estimation). — M.: Mir, 1988. — 264 p. (Russian translation).