CONSTRUCTING STABLE RECURSIVE SCHEMES FOR ESTIMATING PARAMETERS OF STOCHASTIC SYSTEMS

Kirill Chernyshov* and Elena Jharko**

V.A. Trapeznikov Institute of Control Sciences
65 Profsoyuznaya, Moscow 117997, Russia
*E-mail: *myau@ipu.rssi.ru , **zharko@ipu.rssi.ru*

Abstract. A problem of constructing stable recursive algorithms to be used within a broad class of identification and learning problems is considered. An approach is presented leading to obtaining strongly consistent algorithms. Both cases of multi and single input/multi and single output (MIMO, MISO, SISO) linear stochastic dynamic systems are involved. Thus obtained, the recursive algorithms do not involve inversion of the performance index Hessian and are stable to sampled data, in contrast to conventional recursive schemes. Simulation examples are presented, which confirm practical efficiency of the approach.

Key words: Recursive algorithms, Parameter identification, Multi-input/multi-output systems, Colored noise, Condition numbers

1. PROBLEM STATEMENT AND MAIN RESULTS

Consider a strictly causal and asymptotically stable multi input/multi output linear stochastic system with a $n_\Phi \times n_Y$-dimensional (generalized) input process $\Phi(t)$ and a n_Y-dimensional output process $Y(t)$, described by the following input/output relationship

$$Y(t) = \Phi^T(t)\theta^* + V(t) \qquad (1)$$

In (1), θ^* is a n_θ-dimensional ($n_\theta = n_\Phi$) vector of system parameters subject to determination by use current observations of $Y(t)$ and $\Phi(t)$; $V(t)$ is a n_Y-dimensional random process considered as unobservable external disturbances, being the stationary random process having a rational spectral density and described by the relationship

$$V(t) = H_{WY}(q^{-1})W(t)\,, \qquad (2)$$

where $W(t)$ is a n_Y-dimensional stationary white-noise random process, $H_{WY}(q^{-1})$ is the rational matrix transfer function of an asymptotically stable and invertible filter, and q^{-1} is the one step backward shift operator.

System (1) model will be searched for as

$$\hat{Y}(t,\theta) = \Phi^T(t)\theta\,. \qquad (3)$$

By virtue of (3), system (1) output process may be represented in the form

$$Y(t) = \Phi^T(t)\theta + \mathrm{E}(t) \qquad (4)$$

where $\mathrm{E}(t)$ is the equation error (Ljung, 1999, 2002).

Within such a system description, it is assumed that under $H_{WY}(q^{-1}) \equiv 0$, there is unique element in the model set of form (4), having the same input/output description as system (3). Such a condition is equivalent to the following inequality $\mathbf{M}\left\{\Phi_0(t)\Phi_0^T(t)\right\} > 0$ where $\Phi_0(t) \equiv \Phi(t)$ in (1) as $H_{WY}(q^{-1}) \equiv 0$. Here and further, $\mathbf{M}\{\ \}$ stands for the mathematical expectation.

Provided that the filter $H_{WY}(q^{-1})$ in (2) is completely unknown, the only way to obtain a consistent estimate of the parameter vector θ^* of system (1) is using the instrumental variable technique. Within the framework, the most general approach is represented by the extended overdetermined instrumental variable (OIV) method developed by Söderström and Stoica (1989, 2002).

Generically, the parameter vector θ^* estimate corresponding to the extended OIV method may be represented in the form

$$\hat{\theta} = \arg\min_{\theta} I_t(\theta)\ ,\ t \to \infty\ , \tag{5a}$$

$$I_t(\theta) = \frac{1}{t^2}\left(\left\| \sum_{k=1}^{t} Z(k)F(q^{-1})\left(Y(k) - \hat{Y}(k,\theta)\right) \right\|_Q^2 \right). \tag{5b}$$

where $Z(t)$ is the $n_Z \times n_Y$-dimensional instrumental variables matrix, $n_Z \ge \dim\theta$, with both $Y(t)$ and $\hat{Y}(t,\theta)$ being transformed by an asymptotically stable filter $F(q^{-1})$. In (5), a conventional notation $\|X\|_Q^2 = X^T Q X$ is used, where Q is a positively defined weight matrix.

To derive system (1) identification algorithm the following formal representation will be considered

$$\theta(t) = \alpha(t)\theta(t-1) + \Phi(t)\vec{\beta}(t)\ , \tag{6}$$

where $\alpha(t)$ is some scalar coefficient, and $\vec{\beta}(t)$ is some vector-valued coefficient, $\dim\vec{\beta}(t) = n_Y$. These coefficients are to be determined by a condition suitable within the identification problem statement. Such a condition is based on the extended OIV criterion considered above. Namely, substitution of (6) into (5) and taking minimum over $\alpha(t)$ and $\vec{\beta}(t)$ lead to the following criterion

$$\left(\alpha(t), \vec{\beta}(t)\right) = \arg\min_{\alpha,\beta} \frac{1}{t^2} \left\| \sum_{k=1}^{t} Z(k)F(q^{-1})Y(k) - \right.$$
$$\left. - \sum_{k=1}^{t} Z(k)F(q^{-1})\Phi^T(k)\left(\alpha\theta(t-1) + \Phi\vec{\beta}(t)\right) \right\|_Q^2\ , \tag{7}$$

which determines the desired coefficients $\alpha(t)$ и and $\vec{\beta}(t)$.

Thus, representation (6) and criterion (7) imply the following recursive identification algorithm

$$\theta(t) = \theta(t-1) + \Gamma(t)\left(S(t) - R(t)\theta(t-1)\right), \tag{8}$$

$$S(t) = G^T(t)QL(t)\ , \tag{9a}$$

$$R(t) = G^T(t)QG(t)\ , \tag{9b}$$

$$G(t) = G(t-1) + $$
$$+ \frac{1}{t}\left(Z(t)F(q^{-1})\Phi^T(t) - G(t-1)\right), \tag{9c}$$

$$L(t) = L(t-1) + \frac{1}{t}\left(Z(t)F(q^{-1})Y(t) - L(t-1)\right), \tag{9d}$$

where the step $\Gamma(t)$ is determined by the expression

$$\Gamma(t) = \begin{cases} \Lambda(t) + d^{-1}(t)\left(\left(E - \Lambda(t)R(t)\right)\theta(t-1)\theta^T(t-1)\left(E - R(t)\Lambda(t)\right)\right), \\ \qquad \text{as } \det\left[\Phi^T(t)R(t)\Phi(t)\right] > 0 \text{ and } d(t) > 0 \\ \hline \Lambda(t), \text{ as } \det\left[\Phi^T(t)R(t)\Phi(t)\right] > 0 \text{ and } d(t) = 0 \\ \left(\theta^T(t-1)R(t)\theta(t-1)\right)^{-1}\theta(t-1)\theta^T(t-1), \text{ as } \det\left[\Phi^T(t)R(t)\Phi(t)\right] = 0 \end{cases}, \tag{10a}$$

$$d(t) = \theta^T(t-1)R(t)\theta(t-1) - $$
$$\theta^T(t-1)R(t)\Lambda(t)R(t)\theta(t-1)\ ,$$
$$\Lambda(t) = \Phi(t)\left[\Phi^T(t)R(t)\Phi(t)\right]^{-1}\Phi^T(t)\ , \tag{10b}$$

and, in (10), the ratio $\frac{0}{0}$ is considered as 0. Formulae (8)-(10) may de obtained by straightforward calculations.

Thus, the algorithm obtained does not utilizes inversion of the sample identification criterion Hessian, $R(t)$, what reduces sensitivity of the current estimates to variation of the matrix $R(t)$ condition number.

Algorithm (8)-(10) obtained is strongly consistent (i.e. for any initial approximation $\theta(0)$, the recursive sequence of estimates $\{\theta(t)\}$ determined by formulae (8) to (10) converges with probability 1 to θ^* from (1) as $t \to \infty$) as $rank\left(\mathbf{M}\{R(t)\}\right) = n_\theta$. The former condition is known to be valid (*generically*) for a broad class of systems (1) and under an appropriate choice of the instruments $Z(t)$ and the filter $F(q^{-1})$ (Söderström and Stoica, 1989, 2002). The proof of algorithm (8)-(10) consistency is omitted here due to abstract size limitation.

2. EXAMPLES

Below, some examples are presented, which demonstrates convergence properties of the algorithm obtained. Let, for sake of simplicity, both input and output processes be scalar-valued ones, $u(t)$ and $y(t)$ respectively.

Example 1. Let the system be of the form

$$y(t) = -1.41y(t-1) - 0.561y(t-2) -$$
$$-0.019y(t-3) + 0.021y(t-4) + 0.002y(t-5) -$$
$$-0.1u(t-1) - 0.72u(t-2) + v(t) ,$$

where

$$u(t) = \frac{U_{num}\left(q^{-1}\right)}{U_{den}\left(q^{-1}\right)}\omega_i(t) ,$$
$$U_{num}\left(q^{-1}\right) = 0.18q^{-1} - 0.307q^{-2} + 0.049q^{-3} ,$$
$$U_{den}\left(q^{-1}\right) = 1 - 1.468q^{-1} + 0.548q^{-2} -$$
$$-0.061q^{-3} + 0.002q^{-4} ,$$

and $\omega_i(t)$ is the white-noise zero-mean and unit variance Gaussian process,

$$v(t) = \frac{C^*\left(q^{-1}\right)}{D^*\left(q^{-1}\right)}\omega_n(t) ,$$
$$C^*\left(q^{-1}\right) = 0.219q^{-1} - 0.713q^{-2} + 0.015q^{-3} ,$$
$$D^*\left(q^{-1}\right) = 1 - 0.679q^{-1} - 0.059q^{-2} + 0.018q^{-3} +$$
$$+5.944 \cdot 10^{-4} q^{-4} - 9.491 \cdot 10^{-5} q^{-5} ,$$

$\omega_n(t)$ is the white-noise zero-mean Gaussian process, with the standard deviation being equal to 0.05. To form the instrumental variables vector, $z(t)$, the delayed inputs have been chosen, $\left(u(t-1),\ldots,u(t-nz)\right)$, with $nz = 11$. Figure 1 represents behavior of the current identification error norm squares $\eta^2(t) = \left(\theta(t) - \theta^*\right)^T \left(\theta(t) - \theta^*\right)$ corresponding to algorithm (8)-(10) (curve Olc_t, solid line) and recursive extended overdetermined instrumental variables algorithm (curve $Roiv_t$, dotted line), obtained following the methodology of Friedlander (1984). Within the example, $F(q^{-1}) \equiv 1$ and matrix Q has been used as a current estimate of the inverse covariance matrix of the instrumental variable vector, i.e. $Q = \left(\frac{1}{t}\sum_{k=1}^{t}\left(u(k-1),\ldots,u(k-nz)\right)^T \times \left(u(t-1),\ldots,u(t-nz)\right)\right)^{-1}$.

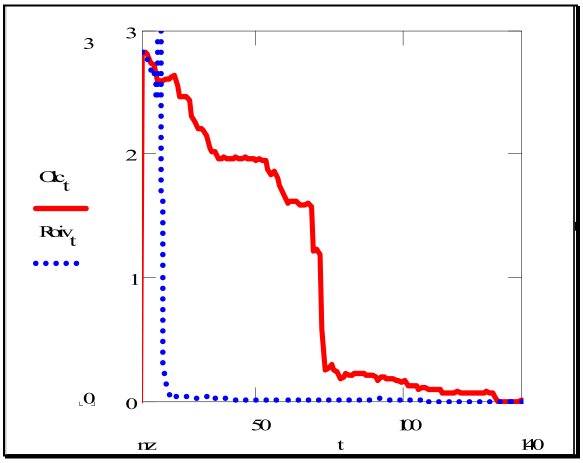

Fig. 1

The condition number of the matrix $R(t)$ from (9), corresponding to such a system is of order 10^2.

Example 2. Consider a model of the same structure as that of example 1 and described as follows

$$y(t) = -0.48y(t-1) + 0.12y(t-2) +$$
$$+0.055y(t-3) - 0.02y(t-4) - 0.001y(t-5) -$$
$$-1.504u(t-1) + 0.213u(t-2) + v(t) ,$$

with the all the other characteristics of both the model and learning algorithms coinciding with those of example 1. Behavior of the values $\eta^2(t)$ for the algorithms is presented on fig. 2a, and fig. 2b represents the behavior in a refined scale.

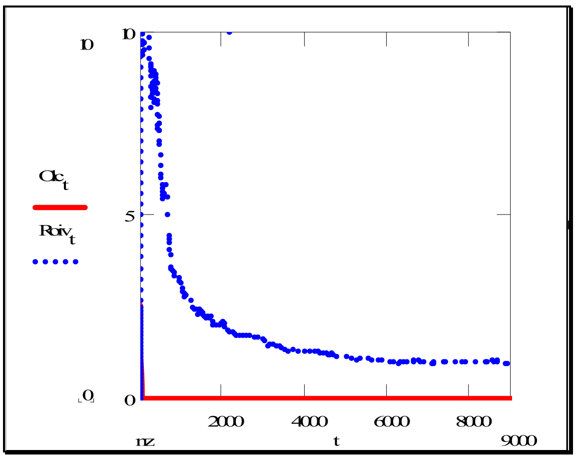

Fig.2a

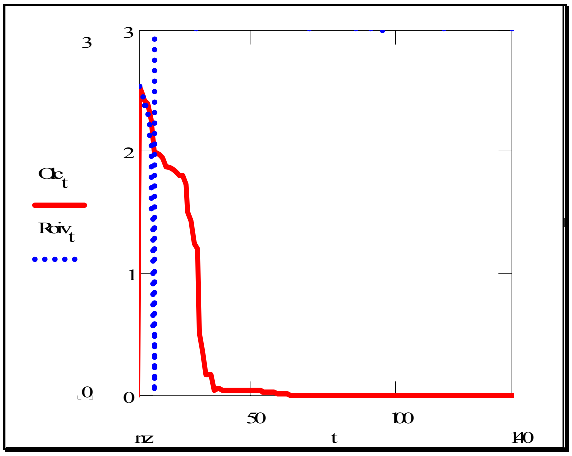

Fig.2b

The condition number of the matrix $R(t)$ from (9), corresponding to such a system is of order 10^4.

Example 3. Let now the disturbance $v(t)$ corresponding to example 2 be the white-noise one, with all the others characteristics being as those of example 2. Corresponding curves are presented at fig. 3.

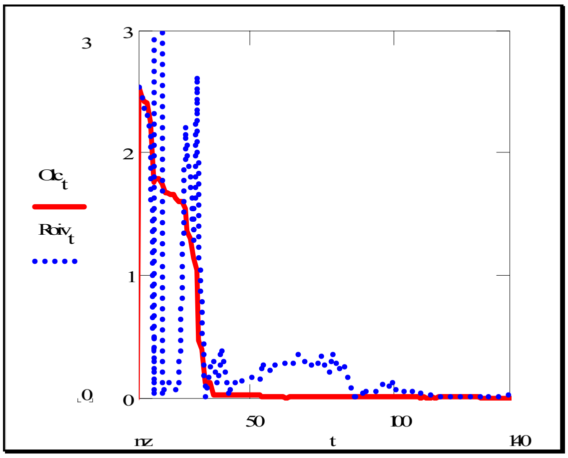

Fig. 3

The condition number of the matrix $R(t)$ from (9), corresponding to such a system is of order 10^4.

Example 4. Let the system subject to identification be of the form

$$y(t) = -0.18y(t-1) - 0.006y(t-2) - \\ -0.009y(t-3) + 0.001y(t-4) - \\ -0.8u(t-1) + 0.3u(t-2) - 0.0346u(t-3) + v(t),$$

where

$$u(t) = \frac{U_{num}\left(q^{-1}\right)}{U_{den}\left(q^{-1}\right)}\omega_i(t),$$

$$U_{num}\left(q^{-1}\right) = -0.1q^{-1} - 0.41q^{-2} + 0.105q^{-3},$$

$$U_{den}\left(q^{-1}\right) = 1 - 1.2q^{-1} + 0.05q^{-2} + \\ +0.222q^{-3} - 0.022q^{-4}$$

and, as above, $\omega_i(t)$ is the white-noise zero-mean and unit variance Gaussian process,

$$v(t) = \frac{C^*\left(q^{-1}\right)}{D^*\left(q^{-1}\right)}\omega_n(t),$$

$$C^*\left(q^{-1}\right) = -0.2q^{-1} - 0.3q^{-2} - 0.144q^{-3},$$

$$D^*\left(q^{-1}\right) = 1 + 1.2q^{-1} - 0.14q^{-2} - \\ -0.588q^{-3} - 0.172q^{-4},$$

$\omega_n(t)$ is the white-noise zero-mean Gaussian process, with the standard deviation being equal to 0.05. All the characteristics of the learning algorithms coincide with those of examples 1 and 2, with dimension of the instrumental variable vector being equal to 15. Behavior of $\eta^2(t)$ for both the algorithms is presented at fig. 4.

The condition number of the matrix $R(t)$ from (9), corresponding to such a system is of order 10^5.

Example 5. Let now the disturbance $v(t)$ from example 4 be a white-noise process, with all the rest characteristics of example to be unchanged. Behavior of $\eta^2(t)$ corresponding to algorithm (8)-(10) is presented at fig. 5 (solid line). Also, at the figure, behavior of $\eta^2(t)$ corresponding to the conventional least-squares (LS) algorithm is demonstrated instead of recursive instrumental variables (dotted line).

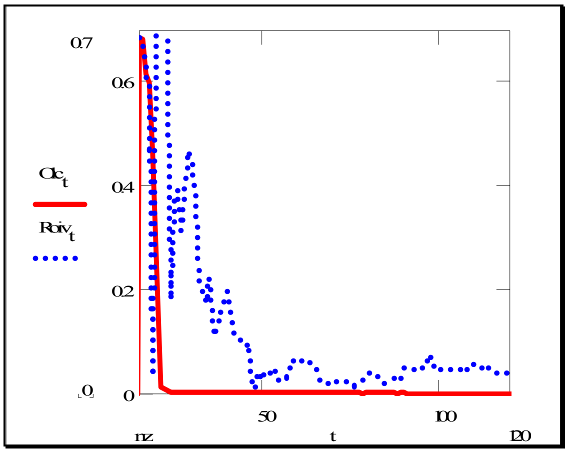

Fig. 4

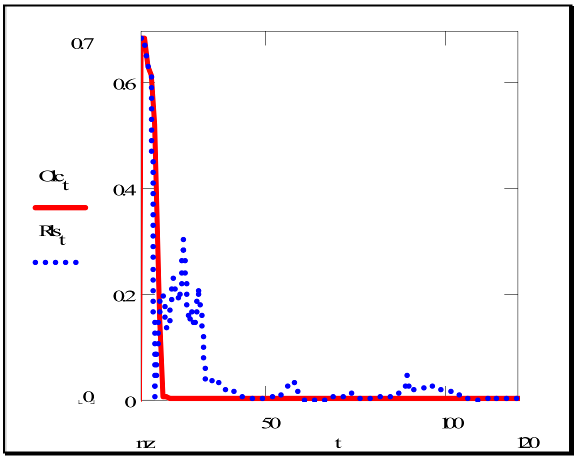

Fig. 5

The condition number of the matrix $R(t)$ corresponding to the system is of order 10^5. The sample covariance matrix of the observation vector used within the ordinary LS-algorithm is of the same order.

Thus, algorithm (8)-(10) demonstrates a good efficiency under various characteristics of systems subject to identification. Stability of the algorithm behavior is clearly manifested both with respect to the external disturbance structure (example 2 and example 3; example 4 and example 5) and to the condition number of the matrix $R(t)$ (example 1 and example 2). From another hand side, the fact, that the identification criterion Hessian is ill-posed, is not necessary an obstacle of convergence of the recursive schemes based on direct minimization of criteria of form like in (5) (examples 3 and 5). Provided that Hessian is ill-posed, just auto-correlation nature of the external disturbances should be considered as a significant issue which considerably affects the sample covariances forming the Hessian $R(t)$ components and, finally, worsening convergence properties of the conventional identification schemes (examples 2 and 4).

3. PARTIAL CASES

Together with general system description (1), behavior convergence of algorithm (8)-(10) within various partial cases is also of interest. In particular, provided that, in (2), $H_{WY}(q^{-1}) \equiv 1$, and imposing simultaneously, in (1), $n_Y = 1$ lead to the following model

$$y(k) = \theta^{*T}\varphi(k) + w(t) \qquad (11)$$

where it was denoted $y(t) = Y(t)$, $\varphi(t) = \Phi(t)$, and $w(t) = W(t)$.

Under system (11), in an analogy with the above considered general case, choosing algorithm (6) coefficients is based on minimization of the conventional criterion $I = \frac{1}{t}\sum_{k=1}^{t}\left(y(k) - \theta^T\varphi(k)\right)^2$ This leads to the following algorithm, which is a partial case of that of (8)-(10) and may be written in the following form

$$\theta(t) = \theta(t-1) + \Gamma(t)\left(\tilde{S}(t) - \tilde{R}(t)\theta(t-1)\right), \quad (12)$$

$$\tilde{S}(t) = \tilde{S}(t-1) + \frac{1}{t}\left(y(t)\varphi(t) - \tilde{S}(t-1)\right), \quad (13a)$$

$$\tilde{R}(t) = \tilde{R}(t-1) + \frac{1}{t}\left(\varphi(t)\varphi^T(t) - \tilde{R}(t-1)\right), \quad (13b)$$

$$\Gamma(t) \overset{def}{=} \begin{cases} d^{-1}(t)\left\{\left(\theta(t-1)\varphi^T(t) - \varphi(t)\theta^T(t-1)\right)\tilde{R}(t)\times \right. \\ \left. \times\left(\varphi(t)\theta^T(t-1) - \theta(t-1)\varphi^T(t)\right)\right\}, & d(t) > 0 \\ \dfrac{\varphi(t)\varphi^T(t)}{\varphi^T(t)\tilde{R}(t)\varphi(t)}, & d(t) = 0 \end{cases}, \quad (14a)$$

$$d(t) = \theta^T(t-1)\tilde{R}(t)\theta(t-1)\varphi^T(t)\tilde{R}(t)\varphi(t) - \left(\theta^T(t-1)\tilde{R}(t)\varphi(t)\right)^2. \quad (14b)$$

Convergence properties of algorithm (12)-(14) are illustrated by the example below. Let, in system (11), the white-noise disturbance $w(t)$ meets the conditions.

$$|w(t)| < r > 0 \ \textit{almost surely}, \quad (15)$$

$$w(t) \ \textit{has symmetric distribution density supported on the interval} \ [-r; r]. \quad (16)$$

Under system description (11), (15), (16), behavior of algorithm (12)-(14) has been compared with that of the “dead-zone” algorithm (Bunich and Bakhtadze, 2003).

$$\theta(t) = \theta(t-1) + \frac{\varphi(t)}{\varphi^T(t)\varphi(t)} f\left(y(k) - \theta^T(t-1)\varphi(k)\right), \quad (17a)$$

$$f(x) = \begin{cases} x - r, & \text{if } x > r \\ 0, & \text{if } x \in [-r; r]. \\ x + r, & \text{if } x < -r \end{cases} \quad (17b)$$

The former is known to possess the property of super-efficiency (Bunich and Bakhtadze, 2003).

Figures 6 and 7 represent behavior of the current identification error norm squares $\eta^2(t) = \left(\theta(t) - \theta^*\right)^T\left(\theta(t) - \theta^*\right)$ corresponding to algorithm (12)-(14) (curve OLC, dark line) and algorithm (17) (curve B83, light line) for a system of (11), (15), (16) having a 30-dimensional parameter vector θ^*, Gausssian distribution of components of the input vector $\varphi(t)$, and uniformly distributed disturbance $w(t)$. Within the example, fig. 6 corresponds to choosing zero initial approximation $\theta(0)$, while fig.

7 corresponds to $\theta(0) = (3,\ldots,3)$. As seen, behavior of algorithm (12)-(14) is asymptotically equivalent to algorithm (17), and mildly enough depends on the initial approximation. Simultaneously, one should be noted that strong consistency conditions of algorithm (12)-(14) are much weaker than those of algorithm (17).

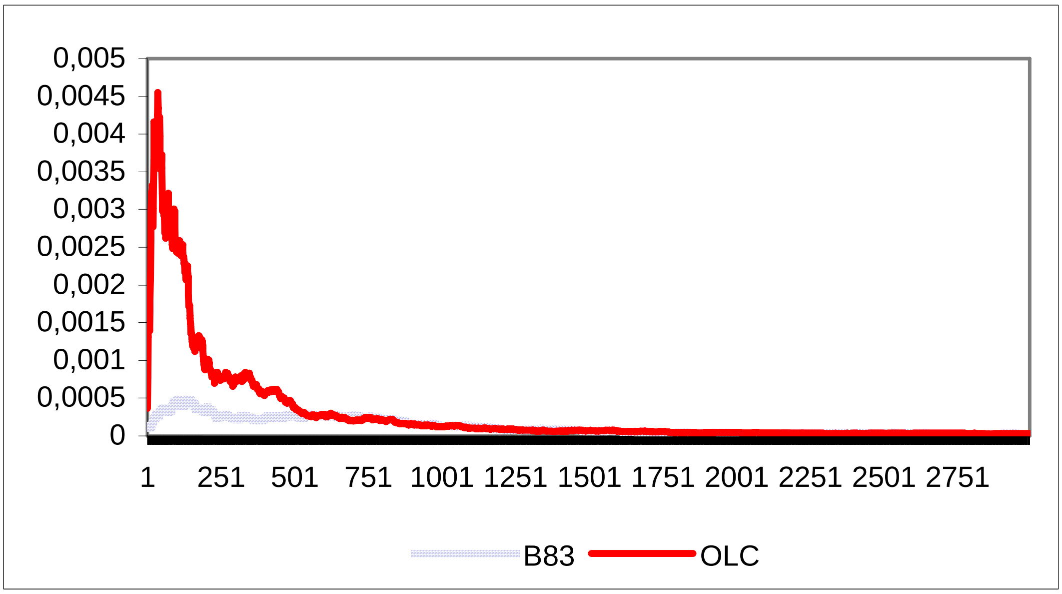

Fig. 6

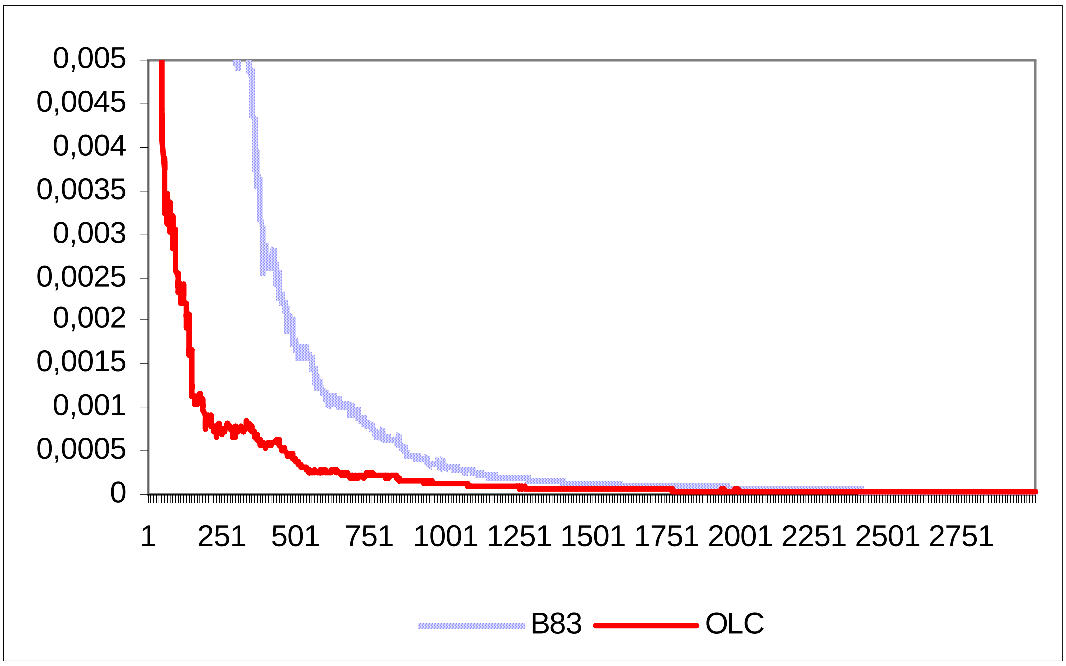

Fig. 7

REFERENCES

Bunich, A.L. and N.N. Bakhtadze (2003). Synthesis and application discrete adaptive control systems with an identifier. Nauka Publ., Moscow, 232 p. ISBN 5-02-006530-7 (in Russian)

Friedlander, B. (1984). "The overdetermined recursive instrumental variable estimation method", *IEEE Trans. Autom. Control*, vol. AC-29, pp. 353-356.

Ljung, L. (1999). System Identification: Theory for the User / 2nd edition. Prentice Hall PRT.

Ljung, L. (2002). "Prediction error estimation methods", Circuits, Systems, and Signal Processing, vol. 21, no. 1, pp. 11-21.

Söderström, T. and P. Stoica (1989). System Identification. Prentice-Hall International, Hemel Hempstead, U.K.

Söderström, T. and P. Stoica (2002). "Instrumental variables methods for system identification", Circuits, Systems, and Signal Processing, vol. 21, no. 1, pp. 1-9.