

# Comparative Analysis of Stochastic Identification Methods and Fault Diagnosis for Multi-mode Discrete Systems<sup>\*</sup>

Olga Fatyanova and Alexey Kondratiev

Ulyanovsk State University, 42 Leo Tolstoy Str., 432970 Ulyanovsk, Russia  
folga@nm.ru, mouse\_fin@mail.ru

**Abstract.** This paper offers the comparative analysis of stochastic approximation methods applied for multi-mode discrete systems identification from incomplete noisy observations. The main item discussed here concerns the rate of convergence of these methods and several aspects that are considered to affect this property. To corroborate the theoretical arguments, the experimental results are provided.

## 1 Introduction

A multi-mode system is understood as a system with changeable coefficients that compounds the vector parameter  $\theta$ . The mode change means that parameter  $\theta$  switches, say, from value  $\theta_1$  to value  $\theta_2$ , taken from a certain compact set of available parameters  $\Theta$ . It is assumed that system is stable if and only if the parameter  $\theta \in \Theta$ . The only data can be used is contained into incomplete noisy observations. So, one can consider two problems to be solved: the problem of mode switching diagnosis, or so-called fault diagnosis, and the problem of identification of the new value  $\theta_2$  of parameter  $\theta$ . We insist on the joint performance of these tasks, using the method of Adaptive Model (AM) and proposing Start-and-Stop Algorithm (SSA), which we. Every time the fault is detected, the SSA launches the identification, which adjusts the AM in order to suit the certain quality requirements. When the identification reaches its goal and demanded agreement with observation is provided, the SSA stops the process. In our research the requirements, mentioned above, are represented in the form of the Auxiliary Functional of Quality (AFQ), compared with certain threshold value. In this article only the identification is discussed, while the description of the SSA can be found in [7].

The outline of the paper is as follows: Section 2 describes the Monitored System (MS) and Kalman filter. In Section 3 we build the AM and the AFQ. Section 4 shows some identification algorithms and certain experimental results are contained in Section 5. Finally, Section 6 concludes the paper.

---

<sup>\*</sup> This work was supported in part by the Russian Ministry of Education (grant No. T02-03.2-3427).

## 2 Monitored System

In the frames of this paper we consider MS without control input with an uncertainty parameter  $\theta$ ,  $\theta \in \mathbb{R}^N$ . First of all, MS includes the object with state equation

$$x(t_{i+1}) = \Phi_\theta x(t_i) + \Gamma w(t_i), \quad i = 0, 1, \dots \quad (1)$$

with  $x \in \mathbb{R}^n$ ,  $\Phi_\theta \in \mathbb{R}^{n \times n}$ ,  $\Gamma \in \mathbb{R}^{n \times q}$  and a noise  $w(t_i)$  where the random initial  $x(t_0)$  at some  $t_0$  has a finite mean  $\bar{x}_0$  and a finite covariance  $P_0 \geq 0$ .

The other part of MS is a sensor with the equation

$$z(t_i) = Hx(t_i) + v(t_i), \quad i = 1, 2, \dots \quad (2)$$

where  $H \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(H) < n$  (incomplete observations) and a noise  $v(t_i)$ . Both  $\{w(t_i)\}$  and  $\{v(t_i)\}$  are i.i.d. zero mean mutually independent wide-sense stationary sequences whose covariances are  $E\{w(t_i)w(t_i)^T\} = Q_\theta \geq 0$  and  $E\{v(t_i)v(t_i)^T\} = R_\theta > 0$  for all  $t_i$ .

The derivation of unbiased estimate  $\tilde{x}(t_i)$  for state  $x(t_i)$  is considered to be the main problem of the project. Assume that the value  $\theta_0$  of parameter  $\theta$  is known in advance and for certain period it remains permanent. Hence in this situation we have *fixed* (fault-free) system mode specified by a nominal value  $\theta_0$  of  $\theta$  and, correspondingly, the steady-state variant of estimation task. This fact allows us to use the classical approach consisting in the use of Kalman filter to tackle our main problem under these simplified conditions.

The standard Kalman filter equations are as follows.

Initial values:

$$\tilde{P}(t_0^+) = P_0, \quad \tilde{x}(t_0^+) = \bar{x}_0 \quad (3)$$

Step 1: time propagation,  $i = 0, 1, \dots$

$$\tilde{x}(t_{i+1}^-) = \Phi_\theta \tilde{x}(t_i^+); \quad (4)$$

$$\tilde{P}(t_{i+1}^-) = \Phi_\theta \tilde{P}(t_i^+) \Phi_\theta^T + \Gamma Q_\theta \Gamma^T. \quad (5)$$

Step 2: measurement update,  $i = 1, 2, \dots$

$$K(t_i) = \tilde{P}(t_i^-) H^T [H \tilde{P}(t_i^-) H^T + R_\theta]^{-1}; \quad (6)$$

$$\tilde{x}(t_i^+) = \tilde{x}(t_i^-) + K(t_i) [z(t_i) - H \tilde{x}(t_i^-)]; \quad (7)$$

$$\tilde{P}(t_i^+) = \tilde{P}(t_i^-) - K(t_i) H \tilde{P}(t_i^-). \quad (8)$$

The MS (1), (2) is assumed to be asymptotically stable in all modes of operating referred to by the subscript  $\theta$ , and so all the processes within the system are wide-sense stationary at every  $t_i$  as  $t_0 \rightarrow -\infty$  (*the main assumption*). The system is designed to hold the *main properties* in all modes:

\*  $(\Phi_\theta, \Gamma Q_\theta^{1/2} \Gamma^T)$  is stabilizable and  $(\Phi_\theta, H)$  is completely observable.

These properties guarantee the existence of the optimal steady-state parameters for a *fixed* (fault-free) system mode

$$\tilde{x}(t_{i+1}^-) = \bar{\Phi}\tilde{x}(t_i^+), \quad \tilde{x}(t_i^+) = \tilde{x}(t_i^-) + \bar{K}[z(t_i) - H\tilde{x}(t_i^-)]$$

with some fixed  $\tilde{x}(t_0^+)$ , constant transition matrix  $\bar{\Phi}$  and constant gain  $\bar{K}$  designed to be optimal in a steady state (as  $t_0 \rightarrow -\infty$ ).

Unfortunately, because of uncertainty classical theory can not be used as it is. In this work we consider two levels of uncertainty.

Case 1: unknown  $Q_\theta, R_\theta$ . The estimate of Kalman gain  $K$  we denote as  $D$ . Here vector  $\theta$  consists of all components of vector  $K$ .

Case 2: unknown  $Q_\theta, R_\theta, \Phi_\theta$ . As in the previous case, we consider  $D$  to be the estimate of  $K$ , and let  $A$  to be the estimate of  $\Phi_\theta$ . Here vector  $\theta$  consists of all components of vector  $K$  and unknown components of matrix  $\Phi_\theta$ .

### 3 Adaptive Model

The problem concerning the derivation of unbiased estimate for  $\Phi_\theta$  in the AM, based on the Kalman filter principle, has a solution only for the case of complete observations [3]. In this work we consider the possibility of evading the strict limitations and accomplishing the task in the case of incomplete observations.

We build AM for MS and Kalman filter with adjustable parameter  $\hat{\theta}$  considered to be the estimate of the corresponding MS parameter  $\theta$ . The property of estimate unbiasedness can be provided only by the optimal choice of the AM structure. In the case of noises with zero mean and gaussian distribution it is the structure of Kalman filter [3]. Let us write the equations of AM:

$$\hat{x}(t_i^-) = A\hat{x}(t_{i-1}^+), \tag{9}$$

$$\hat{z}(t_i) = H\hat{x}(t_i^-), \tag{10}$$

$$r(t_i) = z(t_i) - \hat{z}(t_i) = z(t_i) - H\hat{x}(t_i^-), \tag{11}$$

$$\hat{x}(t_i^+) = \hat{x}(t_i^-) + Dr(t_i). \tag{12}$$

where  $\hat{x}(t_i^\pm)$  is the estimate of  $x(t_i)$ ,  $r(t_i)$  is a difference (residual) between the observation  $z(t_i)$  (immediately obtained from the sensor) and the observation estimate  $\hat{z}(t_i)$  (obtained from the AM).

Hence, the AM allows us to derive the unbiased estimate  $\hat{x}(t_i^\pm)$  for  $x(t_i)$ . Let us consider the initial functional of quality to be built on the base of difference  $e(t_i) = x(t_i) - \hat{x}(t_i)$ . As to [4], error-squared functional of quality is the optimal one for the system with noises having gaussian distribution and zero mean:

$$J_e(t_i) = \frac{1}{2}E[e^T(t_i)e(t_i)], \quad i = 0, 1, \dots \tag{13}$$

While constructing the AFQ, we must take into account the important condition [3]:

- ★ the components of argument  $\hat{\theta}^*$  of AFQ minimization must coincide with corresponding actual values of the entire system parameter (unknown coefficients of  $\Phi_\theta$  in the second case of uncertainty) and optimal value of steady-state filter  $\bar{K}$ .

In practice the value of  $x(t_i)$  is unknown, and our goal is to be find it, therefore AFQ should be built up in another way. Usually the difference  $r(t_i)$  is used for this purpose. So, we take the AFQ in the following form [2]:

$$J_\varepsilon(t_i) = \frac{1}{2} E[\varepsilon^T(t_i)\varepsilon(t_i)], \quad i = 0, 1, \dots \tag{14}$$

where  $\varepsilon(t_i)$  is the history of residuals  $r(t_i)$ , transformed in a special manner to provide

$$J_\varepsilon(t_i) = J_e(t_i) + const, \quad i = 0, 1, \dots \tag{15}$$

Without coming into details of  $\varepsilon(t_i)$  construction, we find it necessary to mention here, that the truth of this condition is provided by the known theorem for the general case [1]. We have fixed the system dimension  $n = 2$  and checked it for this particular case. While proving the fact the constant mentioned above has been determined:  $const = R_\theta$ .

As one can see from (15), the argument of the minimization  $\theta^*$  of  $J_e$  coincides with  $\hat{\theta}^*$ . Hence we will take  $J_e$  as the AFQ, because it is available for practical utilization in the numerical algorithms of identification. Under this point of view the process of identification is evident to turn out the process of optimization or, in other words, of AFQ minimization. The AM is optimal iff (if and only if)

$$\nabla_{\hat{\theta}} J_\varepsilon(t_i) = \varepsilon^T(t_i)S(t_i) = \mathbf{0} \tag{16}$$

where  $\nabla_{\hat{\theta}} J_\varepsilon(t_i) \in \mathbb{R}^{1 \times N}$ , the matrix  $S(t_i) = \nabla_{\hat{\theta}} \varepsilon(t_i)$  is the sensibility matrix [7].

In order to write down  $\varepsilon$  in an explicit form the dimension  $n$  is to be fixed, and below we will operate on 2-order systems in standard observable form:

$$x(t_{i+1}) = \begin{bmatrix} 0 & 1 \\ f_1 & f_2 \end{bmatrix} x(t_i) + \begin{bmatrix} 0 \\ \alpha \end{bmatrix} w(t_i), \quad z(t_i) = [1 \quad 0] x(t_i) + w(t_i).$$

Since  $Q_\theta$ ,  $R_\theta$  and  $\Phi_\theta$  are unknown according to list of uncertainty cases,  $K$  and  $\Phi_\theta$  are to be estimated as  $D$  and  $A$  correspondingly, and almost sure convergence is to be demanded:

$$D = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \rightarrow K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ a_1 & a_2 \end{bmatrix} \rightarrow \Phi_\theta = \begin{bmatrix} 0 & 1 \\ f_1 & f_2 \end{bmatrix}.$$

Now let us write down the equation for  $\varepsilon$  for  $n = 2$ :

$$\varepsilon(t_i) = \mathcal{N}(D)\mathcal{H} = \begin{bmatrix} r(t_{i-1}) \\ d_2 r(t_{i-1}) + r(t_i) \end{bmatrix} = \begin{bmatrix} \varepsilon_1(t_i) \\ \varepsilon_2(t_i) \end{bmatrix}.$$

$$\mathcal{N}(D) = \begin{bmatrix} 1 & 0 \\ d_2 & 1 \end{bmatrix}, \quad \mathcal{H} = \begin{bmatrix} r(t_{i-1}) \\ r(t_i) \end{bmatrix}.$$

To obtain the gradient of AFQ (for the first case of uncertainty) the sensibility matrix should be calculated.

$$S_D(t_i)^T = \begin{bmatrix} \frac{\partial \varepsilon_1(t_i)}{\partial d_1} & \frac{\partial \varepsilon_2(t_i)}{\partial d_1} \\ \frac{\partial \varepsilon_1(t_i)}{\partial d_2} & \frac{\partial \varepsilon_2(t_i)}{\partial d_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial r(t_{i-1})}{\partial d_1} & d_2 \frac{\partial r(t_{i-1})}{\partial d_1} + \frac{\partial r(t_i)}{\partial d_1} \\ \frac{\partial r(t_{i-1})}{\partial d_2} & r(t_{i-1}) + d_2 \frac{\partial r(t_{i-1})}{\partial d_2} + \frac{\partial r(t_i)}{\partial d_2} \end{bmatrix}$$

For this purpose we differentiate componentwise the equations of AM by  $D$ , taking in account(4), (5), (6), (7), (8).

$$\frac{\partial \hat{x}(t_i^-)}{\partial d_1} = A \frac{\partial \hat{x}(t_{i-1}^+)}{\partial d_1}, \quad \frac{\partial r(t_i)}{\partial d_1} = - [1 \quad 0] \frac{\partial \hat{x}(t_i^-)}{\partial d_1},$$

$$\frac{\partial \hat{x}(t_i^+)}{\partial d_1} = \frac{\partial \hat{x}(t_i^-)}{\partial d_1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} r(t_i) + D \frac{\partial r(t_i)}{\partial d_1}.$$

$$\frac{\partial \hat{x}(t_i^-)}{\partial d_2} = A \frac{\partial \hat{x}(t_{i-1}^+)}{\partial d_2}, \quad \frac{\partial r(t_i)}{\partial d_2} = - [1 \quad 0] \frac{\partial \hat{x}(t_i^-)}{\partial d_2},$$

$$\frac{\partial \hat{x}(t_i^+)}{\partial d_2} = \frac{\partial \hat{x}(t_i^-)}{\partial d_2} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} r(t_i) + D \frac{\partial r(t_i)}{\partial d_2}.$$

Hence,  $\frac{\partial \hat{x}(t_i^-)}{\partial d_1}$ ,  $\frac{\partial r(t_i)}{\partial d_1}$ ,  $\frac{\partial \hat{x}(t_i^+)}{\partial d_1}$ ,  $\frac{\partial \hat{x}(t_i^-)}{\partial d_2}$ ,  $\frac{\partial r(t_i)}{\partial d_2}$ ,  $\frac{\partial \hat{x}(t_i^+)}{\partial d_2}$  are calculated iteratively and then substituted into the  $S_D(t_i)^T$  matrix. The initial values is necessary to set:

$$\frac{\partial \hat{x}(t_0^+)}{\partial d_1} = 0, \quad \frac{\partial \hat{x}(t_0^+)}{\partial d_2} = 0.$$

For second case of uncertainty the sensibility model is designed analogically.

## 4 Stochastic Approximation

In this paper the process of identification is performed by discrete searchless algorithms, which are also called the methods of stochastic approximation and represented below in the general recurrent form [5]:

$$\hat{\theta}[t_j] = \hat{\theta}[t_{j-1}] - A[t_j] S^T[t_{j-1}] \varepsilon[t_{j-1}]. \tag{17}$$

The choice of  $A[t_j]$  affects the rate of convergence greatly. The sufficient conditions of algorithm convergence can be found in [5], [6]. The identification method, corresponding diagonal form of  $A[t_j]$  is discussed in 4.3. The case of equal diagonal components  $\lambda[t_j]$  see below in 4.1.

### 4.1 Algorithm 1. Robbins-Monro Algorithm

Let us take  $\lambda[t_j] = \frac{1}{j}$ . By doing so we get the algorithm representing the stochastic analogue of gradient method. This method is also called the Robbins-Monro multidimensional procedure of stochastic approximation. [6], [7]:

$$\hat{\theta}[t_{j+1}] = \hat{\theta}[t_j] - \lambda[t_{j+1}] S^T[t_j] \varepsilon[t_j], \quad \lambda[t_{j+1}] = \frac{1}{j+1}, \quad j = 1, 2, \dots \tag{18}$$

## 4.2 Algorithm 2. Optimal Algorithm

The information about the class of density of distribution the noises belong is of primary importance, because it helps to make an optimal choice of matrix  $\Lambda$ . The algorithm, corresponding to this optimal choice provides the best rate of convergence of  $\hat{\theta}[t_j]$  to  $\hat{\theta}^*$  among all methods [4]. So, for the case of noises having gaussian distribution with zero mean and unknown covariances least-squares method (LSM) is the optimal one. To derive LSM the general form of identification algorithm is to be written down:

$$\hat{\theta}[t_{j+1}] = \hat{\theta}[t_j] - \Lambda[t_{j+1}]S^T[t_j]\varepsilon[t_j], \quad j = 1, 2, \dots \quad (19)$$

where  $\Lambda[t_j] \in \mathbb{R}^{N \times N}$ . In LSM  $\Lambda[t_j]$  is determined by the formula [4]:

$$\Lambda[t_{j+1}] = \Lambda[t_j] - \Lambda^{-1}[t_j]S^T[t_j](S[t_j]\Lambda[t_j]S^T[t_j])^{-1}S[t_j]\Lambda[t_j]. \quad (20)$$

The equation of approximation written down through the increment is

$$\hat{\theta}[t_{j+1}] = \hat{\theta}[t_j] + \Delta\hat{\theta}[t_j], \quad j = 1, 2, \dots \quad (21)$$

We denote  $P[t_j] = \Lambda^{-1}[t_j]$  and from (19), (20), (21) obtain the LSM:

- set the initial value  $P[t_1] = I$ ;
- calculate  $P[t_{j+1}] = P[t_j] + S[t_j]^T S[t_j]$ ;
- find  $\Delta\hat{\theta}[t_j]$  from the system of linear equations  $P[t_{j+1}]\Delta\hat{\theta}[t_j] = -S^T[t_j]\varepsilon[t_j]$
- calculate new value of parameter  $\hat{\theta}[t_{j+1}] = \hat{\theta}[t_j] + \Delta\hat{\theta}[t_j]$

The disadvantage of the LSM is that for each step the solution of the system of linear equations is sought. This operation is quite laborious to perform, so it is worth to modify the method in order to avoid it. Certainly the newly obtained algorithm is not optimal in the strict sense, that is why we call it suboptimal, nevertheless its employment let us to cut down computational costs in comparison with the LSM. In the later section suboptimal method is discussed in details.

## 4.3 Algorithm 3. Suboptimal Algorithm

So-called suboptimal algorithm is developed on the base of optimal one by transforming the  $S^T[t_j]S[t_j]$  matrix to diagonal form. Only diagonal elements of matrix product  $S^T[t_j]S[t_j]$  remain unchanged, the rest are substituted for zeros:

$$S^T[t_j]S[t_j] \approx \begin{bmatrix} \left\| \frac{\partial \varepsilon[t_j]}{\partial \theta_1} \right\|^2 & \dots & 0 \\ \dots & \ddots & \vdots \\ 0 & \dots & \left\| \frac{\partial \varepsilon[t_j]}{\partial \theta_N} \right\|^2 \end{bmatrix} \quad (22)$$

Thus, we have got the suboptimal algorithm:

- set the initial value  $P[t_1] = I$ ;
- calculate diagonal elements of matrix  $P$ ,  $p_{ii}[t_{j+1}] = p_{ii}[t_j] + \left\| \frac{\partial \varepsilon[t_j]}{\partial \theta_i} \right\|^2$ , for  $i = 1, 2, \dots, N$
- calculate new value of parameter  $\hat{\theta}_i[t_{j+1}] = \hat{\theta}_i[t_j] - \frac{S^T[t_j]\varepsilon[t_j]}{p_{ii}[t_{j+1}]}$ ,  $i = 1, 2, \dots, N$

#### 4.4 Stability Requirements. Jury's Criteria

From the condition of AM stability  $\rho[(I - DH)A] < 1$  the characteristic equation can be obtained [2]:

$$q(\lambda) = b_0\lambda^2 + b_1\lambda + b_2 = 0 \tag{23}$$

where  $b_0 = 1$ ,  $b_1 = d_2 - a_2$ ,  $b_2 = a_1(d_1 - 1)$ . The stability condition according to Jury criteria may be written as [2]:

$$q(1) > 0, \quad (-1)^N q(-1) > 0, \quad b_0 > |b_N|. \tag{24}$$

In particular case of 2-order system these conditions transform to the next ones:

$$1 + b_1 + b_2 > 0, \quad 1_0 - b_1 + b_2 > 0, \quad 1 > |b_2|. \tag{25}$$

Hence, we have got an opportunity to determine the stability of a system at every step of approximation process and to work out the heuristic algorithm, that would be capable of preventing the AM crash.

The main idea of this method is to keep the parameter  $\hat{\theta}(t_{j+1})$  within the set of stability  $\Theta$ . Therefore we have designed the algorithm reasoning from the idea that identification is to be reinitiated, if the AM became unstable, and launched with a new stable value.

- calculate  $\hat{\theta}(t_{j+1})$ ;
- check Jury's criteria (24); if the AM with parameter  $\hat{\theta}(t_{j+1})$  is stable, continue identification; else choose some  $\hat{\theta}_{new}(t_{j+1}) \in \Theta$ .

Now the question is: how to find  $\Theta$  and choose  $\hat{\theta}_{new}$  during the approximation? Let us consider this problem for 2-order system and write down the conditions of stability for  $d_1$  and  $d_2$ :

$$1 - \frac{1}{|a_1|} < d_1 < \frac{1}{|a_1|} + 1, \tag{26}$$

$$-(a_1(d_1 - 1) + 1) + a_2 < d_2 < (a_1(d_1 - 1) + 1) + a_2. \tag{27}$$

From (26) it is reasonably to take  $d_1 = 1$  and  $d_2 = a_2$  from (27). In equivalent manner we take  $a_1 = 0$  and  $a_2 = d_2$ . Thus, it is guaranteed that  $\hat{\theta}_{new}(t_{j+1}) \in \Theta$ .

## 5 Experiments

The workability of the proposed ideas is to be approved by the experiments for 2-order system. The subject of our interest is the rate of convergence. We consider several aspects to affect this characteristic of identification: the method used in adaptation process, the case of uncertainty, the level of noise both in object and sensor ( $Q_\theta$  and  $R_\theta$ ), and the stability of entire system (eigenvalues of matrix  $\Phi_\theta$  from the Jury's criteria point of view).

In real life the parameter  $\theta$  is beyond our reach, while in the experimental conditions our programm simulates both entire system and AM, and we know  $\theta$  (we, but not the AM!) and are able to compare it with  $\hat{\theta}$ . Hence we have “a task with known solution”.

Therefore the scheme of all experiments is as follows (unless otherwise arranged): during the first thousand of iteration  $\theta = \hat{\theta}$ , then at  $i = 1000$  we simulate a fault by changing  $\hat{\theta}$  and at  $i = 1000$  the process of approximation is launched. In this type of experiments the SSA is not used.

Let us consider the first case of uncertainty and take MS with

$$\Phi_{\theta} = \begin{bmatrix} 0 & 1 \\ 0.30 & 0.67 \end{bmatrix}.$$

which eigenvalues (approximately 0.307 and 0.977) are close to stability limits. The table below contains the convergence time (in iterations) which is averaged over 100 realization of approximation process. The 15-percent ratio error is taken (see fig. 1, fig. 2 and fig. 3).

		Alg. 1.		Alg. 2		Alg. 3	
$Q$	$R$	$d_1$	$d_2$	$d_1$	$d_2$	$d_1$	$d_2$
1	1	14479	4675	5967	599	6698	963
1	0.1	17766	651	4314	493	4716	1138

Now the MS with eigenvalues which equal approximately -0.358 and 0.558 and, hence, are far from stability limits:

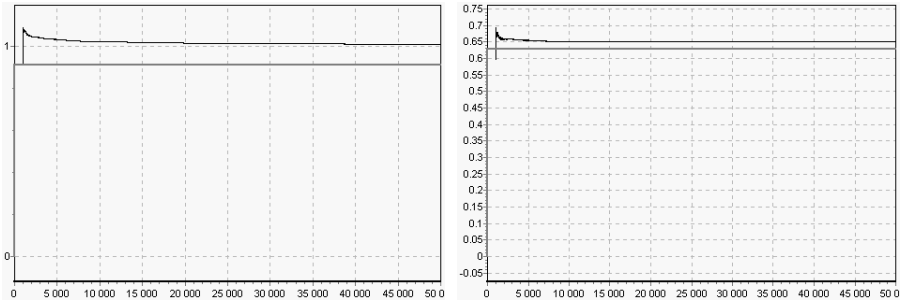
$$\Phi_{\theta} = \begin{bmatrix} 0 & 1 \\ 0.20 & 0.20 \end{bmatrix}.$$

The average convergence time is represented in the table:

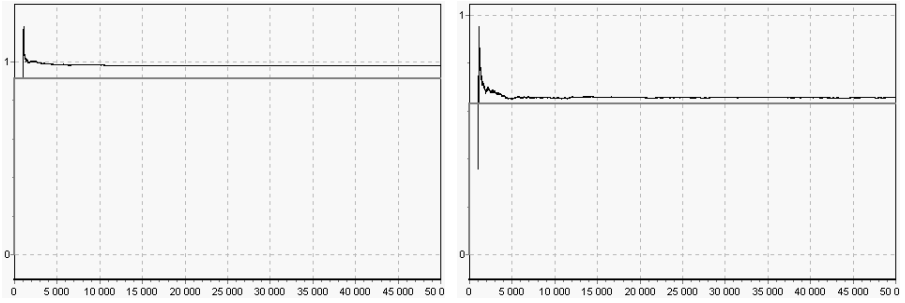
		Alg. 1.		Alg. 2		Alg. 3	
$Q$	$R$	$d_1$	$d_2$	$d_1$	$d_2$	$d_1$	$d_2$
1	1	38376	39822	13626	18380	11928	21387
1	0.1	16889	8440	4431	8610	5566	8000

In the second case of uncertainty the convergence of  $A$  is much slower than the one of  $D$ . To explain this fact series of experiments were made. The parameter  $D$  was taken coincided with  $\bar{K}$  and fixed. The AFQ was considered as a function of  $a_1$  and  $a_2$ , and level lines, consisting of the points with coordinates  $(a_1, a_2) : J_{\varepsilon}(a_1, a_2) = const$ . These lines look like oblong ellipses. It is typical for ravine surfaces, so one take  $f_1 = 0.2$ ,  $f_2 = 0.7$  and in the first realization of approximation process get  $a_1[t_{max}] \approx 0.7$  and  $a_2[t_{max}] \approx 0.2$ , in the second  $a_1[t_{max}] \approx 0.4$  and  $a_2[t_{max}] \approx 0.5$ , then  $a_1[t_{max}] \approx 0.8$  and  $a_2[t_{max}] \approx 0.1$ , or even  $a_1[t_{max}] \approx -1.3$  and  $a_2[t_{max}] \approx 0.4$ . The fact that the “correct answer” is also available, is very important, because for all these realizations the error  $e(t_{max})$  does not vary greatly, as well as  $J_{\varepsilon}(a_1, a_2)$ , so  $\hat{x}$  is “good enough”. Taking these features in account one may note that  $a_1(t_{max}) + a_2(t_{max}) \approx 0.9 = const$ , as if points  $(a_1[t_{max}], a_2[t_{max}])$  belong to certain line or oblong ellipse.

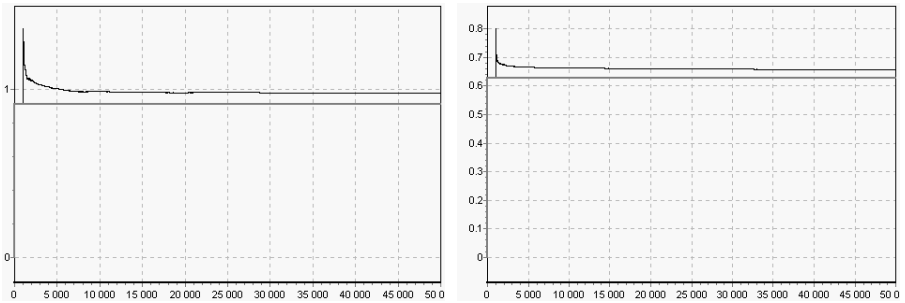




**Fig. 1.** Tracking the parameters  $k_1$  and  $k_2$  by  $d_1$  (left) and  $d_2$  (right) for Algorithm 1



**Fig. 2.** Tracking the parameters  $k_1$  and  $k_2$  by  $d_1$  (left) and  $d_2$  (right) for Algorithm 2



**Fig. 3.** Tracking the parameters  $k_1$  and  $k_2$  by  $d_1$  (left) and  $d_2$  (right) for Algorithm 3

### 5.1 Experimental Conclusions

- The ravine character of AFQ (as a function of  $a_1$  and  $a_2$ ) degrades seriously the rate of convergence, that does not allow to get correct estimates of  $\Phi_\theta$  for appropriate time. The enlargement of modelling time to 500000 iterations gives nothing, the millions of iterations are demanded. Nevertheless  $x(t_i)$  is estimated good enough to say that the main task – to derive  $\hat{x}(t_i)$  – is solved.

- The series of experiments approves that LSM has got the best rate of convergence. Suboptimal method considered to be the successful improvement of LSM, because this algorithm provides serious computational shortcut and insignificantly yields to LSM in the rate of convergence.
- The larger  $R$  (in comparison with  $Q$ ) the worse the rate of convergence.
- The rate of convergence is better for MS which eigenvalues are close to the limit of stability. It is easily seen from (1): such MS possess more strongly marked inner dynamics than another ones.

## 6 Conclusions and Future Work

Basing on the results of the experiments we can say that identification methods can be successfully applied for the solution of the main task, the derivation of unbiased estimate of  $x$ . Unfortunately we face the challenge in obtaining correct estimations in the second case of uncertainty, because of the ravine character of AFQ. So one of the directions of the future works is to minimize the influence of this factor to the rate of convergence. Another direction is to study new identification methods such as ones of Newton's type and conjugated gradients algorithm.

## References

1. Semoushin, I.V., Tsyganova, J.V.: Indirect error control for adaptive filtering. In: Neittaanmaki, P., Tiihonen, T., Tarvainen, P. (eds.): Proc. of the 3rd European Conference on Numerical Mathematics and Advanced Applications. World Scientific, Singapore New Jersey London Hong Kong (2000) 333–340
2. Semoushin, I.V.: Adaptive Identification and Fault Detection Methods in Random Signal Processing. Saratov University Publishers, Saratov (1985) [*in Russian*]
3. Semoushin, I.V.: Identification of linear stochastic objects on the base of incomplete noisy observations. Automatics and telemekhanics, (1985) 61–71 [*in Russian*]
4. Tsypkin, J.Z.: Optimal identification of dynamic objects. Mensuration control automation, Vol. 3 (1983) 47–59 [*in Russian*]
5. Tsypkin, J.Z.: Adaptation and training in automatic systems. Moscow (1968) [*in Russian*]
6. Vasan M.: Stochastic approximation. Moscow (1972) [*in Russian*]
7. Ponyrko S.A., Semoushin, I.V.: About the choice of start and stop algorithm for minimisation of root-mean-square quality criterion. Autometria, Vol. 2 (1973) 68–74 [*in Russian*]