American Journal of Engineering Research (AJER) ***2022***
American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-11, Issue-03, pp-41-51
www.ajer.org
Research Paper Open Access

Robust Identification of Dual – Rate Systems Based on Accelerated Stochastic Approximation and Bayesian Information criterion

Vojislav Filipovic
Department for Automatic Control, Robotic and Fluid Technique, Faculty of Mechanical and Civil Engineering, University of Kragujevac, Dositejeva 19, 36000 Kraljevo, Serbia

ABSTRACT: *In the paper is proposed outlier robust identification algorithm for dual – rate stochastic systems. The algorithm is accelerated stochastic approximation which is based on averaging in both iterates and observations. In comparison with algorithm based only on averaging in iterates, proposed in the paper algorithm, is more stable in the initial period. The derivation of the algorithm is based on modified Newton – Raphson algorithm. In the paper is considered identification of fast rate model without any transformation. It is used output-error model philosophy. The auxiliary model is FIR (finite impulse response) model. The identification consists from two stages. In the first stage it is estimated FIR model. The parameter estimation of FIR model is performed by using outliers robust accelerated stochastic approximation and order of the model is determined with outliers robust Bayesian information criterion. In the second stage, by using robust accelerated stochastic approximation is estimated fast rate model. The main contributions of the paper are: (i) Design of the robust accelerated stochastic approximation algorithm based on modified Huber theory; (ii}) design of robust Bayesian information criterion.*

KEYWORDS*: Dual – rate system· Averaging in the iterates· Averaging in the observations· Stochastic approximation· Robust Bayesian information criterion.*

Date of Submission: 02-03-2022 Date of acceptance: 17-03-2022

I. INTRODUCTION

Since 1950s multi - rate systems (systems with different inputs and outputs sampling periods) have attracted significant attention [1-2]. Such processes exist in process industry [3-4]. The cascade control system is multi – rate system [5]. Here the inner loop is a fast process and the outer loop is slow one. The sampling period of the signals in the inner loop is small and sampling period of the signals in the outer loop is relatively large. The special class of multi – rate systems is a dual – rate systems and last kind of system will be considered in this paper. Networked control with multi – rate sampling is considered in reference [6]. Stabilization problem of the multi – rate networked control system is transformed to the stabilization problem of the lifted feedback system. The reference [7] describes identification of dual – rate ARX based on filter – based stochastic gradient algorithm.

The key field in control theory is system identification. The most of identification algorithms assume that input and output signals have same sampling periods. For dual – rate systems these algorithms cannot be applied directly. The development of model structure that is suitable for different input – output sampling rate is the first task for parameter estimation. Two main methods to transform the dual – rate model are: (i) lifting technique in the area of state space models [8]; (ii) polynomial transformation technique which can derive a dual – rate model that directly utilize all available data (fast input and slow output data) [9]. Lifted and transformed models contain more parameters than the original systems. The drawback of these methods is that they require identifying more parameters.

The output – error method [10] is based on information vector which includes output of the model instead output of the system. Such procedure represents the basis for identification of the fast model in identification of dual – rate systems. The idea is exploited in references [11-12]. In the paper [11] is first estimated finite impulse response model (FIR) from dual – rate data. Using FIR model one can predict the noise – free fast rate outputs. Finally, a fast single – rate model is identified. That is ideological basis for our paper.

The original stochastic approximation algorithm [15] is simple for implementation but has slow convergence rate. The main research effort is to improve speed of convergence of stochastic approximation. The significant step in such directions is made by Polyak [14]. That is now well known averaging of the iterates [15]. Here the gain in the stochastic approximation is selected to be larger than gain in classical procedure [13]. Further improvement is based on simultaneously average iterates as well as observations [16]. So is given accelerated stochastic approximation.

It is well known that the structure identification of the model can be done by information criterions. Many information criterions exist. Very often the Bayesian information criterion (BIC) is used [17-18].

In this paper we considered case when population of observations include outliers. The main mathematical machinery for such case is a robust statistics [19]. In that case the main tool for design of identification algorithms is a Huber`s loss function which depends from least favourable probability density for disturbance. This function is only first order differentiable making it impossible to use Newton – Raphson method. To overcome the above problem in this paper is introduced smooth version of the Huber`s loss function (pseudo Huber`s loss function) which have derivatives of all degrees [20-21]. In this paper is proposed Newton – Raphson procedure which includes observations averaging [16]. In the algorithm for first derivative the modified Huber`s loss function (includes observation averaging) is used and for second derivative the modified pseudo – Huber`s loss function is used. The given algorithm is extended with averaging in iterates. So is given algorithm with averaging in both iterates and observations.

With above algorithm and robust BIC criterion (based on Huber's loss function) it is formed finite number of alternatives which depend from parameter and dimension of parameter. Minimum of BIC criterion among all alternatives gives estimate of parameter and order of FIR model. The FIR model is used for calculation of noise free outputs. That is the first stage. The second stage is used accelerated robust stochastic approximation algorithm for identification of fast model of considered output-error model.

The main contributions of the paper are

(i) Outlier robust two stage identification of fast model of dual – rate systems.

(ii) The first stage (identification of FIR model from dual – rate data) consists of robust accelerated algorithm, based on averaging in both iterates and observations, and new robust version of BIC criterion

(iii) By using the estimated FIR model, from first stage, it is possible to predict noise free fast rate output and, in the frame of output-error philosophy, perform identification of fast rate model with robust accelerated stochastic approximation algorithm.

II. SOME ACCELERATED STOCHASTIC APPROXIMATION ALGORITHM FOR SYSTEM IDENTIFICATION

Stochastic approximation algorithm [13] is a very important algorithm for application in different scientific fields (control, signal processing). The main drawback of the algorithm is its slow convergence. A lot of efforts have been put into improvement of rate of convergence of algorithm [22]. Significant step is made by Polyak [14]. In what follows we review some accelerated algorithms. For all of them we consider next regression model

$$y(k) = \boldsymbol{\varphi}^T(k)\boldsymbol{\theta} + e(k) \qquad (1)$$

where $\varphi(k) \in R^m$ is a vector of inputs, $y(k) \in R^1$ is output, $e(k) \in R^1$ is a stochastic disturbance and $\theta \in R^m$ is a vector of parameters. Standard stochastic approximation has a form

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + a(k)\boldsymbol{\varphi}(k)\left[y(k) - \boldsymbol{\varphi}^T(k)\hat{\boldsymbol{\theta}}(k-1)\right]$$

(2) For a convergence of procedure (2) the gain $a(k)$ satisfies next condition

$$a(k) \to 0 \;,\; \sum_{k=1}^{\infty} a(k) = \infty \;,\; \sum_{k=1}^{\infty} a^2(k) < \infty, \qquad (3)$$

The common choice for $a(k)$ is $a(k) = 1/k$. We further will consider accelerated stochastic approximation algorithms.

ALGORITHM I: Averaging in the iterates [14]

The idea here is to generate a sequence of rough estimate using higher gain $a(k)$ first and then take arithmetic averages of the resulting iterates. The algorithm has a form

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + a(k)\boldsymbol{\varphi}(k)\left[y(k) - \boldsymbol{\varphi}^T(k)\hat{\boldsymbol{\theta}}(k-1)\right] \qquad (4)$$

$$\bar{\boldsymbol{\theta}}(k) = \frac{1}{k}\sum_{i=1}^{k}\hat{\boldsymbol{\theta}}(i) \qquad (5)$$

$$\frac{a(k)}{a(k+1)} = 1 + o(a(k)) \;,\; a(k) \to 0 \;,\; \sum_{i=1}^{\infty} a(k) = \infty \qquad (6)$$

The common choice of gain $a(k)$ is

$$a(k) = \frac{1}{n^{\gamma}} \;,\; \frac{1}{2} < \gamma < 1 \qquad (7)$$

In reference [14] it is supposed that stochastic disturbance $e(\cdot)$ is independent. The more general case of disturbance is considered in reference [15].

ALGORITHM II: Averaging in iterates and feedback [24]

The algorithm (4) – (5) is "off– line" in the sense that value $\bar{\boldsymbol{\theta}}(k)$ is not used in relation (4). To overcome that problem in reference [24] is proposed next algorithm.

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + a(k)\boldsymbol{\varphi}(k)\left[y(k) - \boldsymbol{\varphi}^{T}(k)\hat{\boldsymbol{\theta}}(k-1)\right] + \\ + a(k)A(\bar{\boldsymbol{\theta}}(k-1) - \hat{\boldsymbol{\theta}}(k-1)) \;,\; A > 0 \qquad (8)$$

where $a(k)$ and $\bar{\boldsymbol{\theta}}(k)$ are given with relations (7) and (5).

That modification improves rate of convergence of stochastic approximation algorithm.

ALGORITHM III: Averaging in both iterates and observations [16]

In the Algorithm I and Algorithm II the averaging should be carried out after the iterations have passed the transient period [14]. Namely, taking averages in the first a few iterations may result in poor performance and create large errors. Owing that fact to improve initial performance of the algorithm is very important and algorithm with average in both iterates and observations appear to be more stable in the initial period. The algorithm has a form

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + a(k)\sum_{i=1}^{k}\boldsymbol{\varphi}(i)\left[y(i) - \varphi^{T}(i)\hat{\boldsymbol{\theta}}(i-1)\right] \qquad (9)$$

$$\bar{\boldsymbol{\theta}}(k) = \frac{1}{k}\sum_{i=1}^{k}\hat{\boldsymbol{\theta}}(k) \qquad (10)$$

where $a(k)$ is given with relation (7).

III. ROBUST ACCELERATED STOCHASTIC APPROXIMATION ALGORITHM

In this part of the paper we define the two stage procedure for identification of dual – rate models. The main goal is to identify a fast single – rate model. Let us suppose that discrete model has a next form

$$y(k) = x(k) + e(k) \qquad (11)$$

where $y(k) \in R^{1}$ is output of the system, $e(k)$ is stochastic disturbance and $x(k)$ is given by next relation

$$x(k) = P\left(q^{-1}\right)u(k) \qquad (12)$$

where $y(k) \in R^{1}$ is input signal and q^{-1} is time delay operator $q^{-1}x(k) = x(k-1)$. In relation (12) $P\left(q^{-1}\right)$ is the discrete transfer function of the corresponding single – rate system. The function $P\left(q^{-1}\right)$ has a form

$$P\left(q^{-1}\right) = \frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)} \qquad (13)$$

$$B\left(q^{-1}\right) = b_1 q^{-1} + \ldots + b_m q^{-m} \;,\; A\left(q^{-1}\right) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n}$$

Let us suppose that available input – output data are $\{u(k)\ ,\ k=0,1,2...\}$ at the fast rate and $\{y(k\Delta)\ ,\ k=0,1,2...\}\ ,\ \Delta \geq 2$ is integer, for slow rate. The inter sample outputs are $y(k\Delta+i)\ ,\ i=0,1,2...\ ,\ \Delta-1$ not available.

Now we will present two stage identification procedure for fast rate model.

First stage (identification of FIR model)

Here we find the equivalent FIR model for next model

$$y(k)=\frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k)+e(k) \qquad (14)$$

Let us assume that finite impulse response model $G\left(q^{-1}\right)$ can be used as equivalent presentation of model (14)

$$y(k)=G\left(q^{-1}\right)u(k)+e(k) \qquad (15)$$

where

$$G\left(q^{-1}\right)=g_1q^{-1}+g_2q^{-2}+...+g_pq^{-p} \qquad (16)$$

The vector form of the model (15) is

$$y(k)=\boldsymbol{\varphi}^T(k)\boldsymbol{\theta}+e(k) \qquad (17)$$

where

$$\boldsymbol{\theta}=\left[g_1,g_1,...,g_p\right]^T\in R^p$$

$$\boldsymbol{\varphi}(k)=\left[u(k-1),u(k-2),...u(k-p)\right]^T\in R^p$$

The identification of $\boldsymbol{\theta}$ in model (17) in the dual – rate setting is possible if we replace k with $k\Delta$ and model (17) becomes.

$$y(k\Delta)=\boldsymbol{\varphi}^T(k\Delta)\boldsymbol{\theta}+e(k\Delta) \qquad (18)$$

It is supposed that population of observations contains the outliers. In such case relevant assumption that a priori known class of distributions to which belongs disturbance. The class of distributions is

$$P_\varepsilon=\left\{P:P=\left(1-\varepsilon^*\right)N_D+\varepsilon^*G\ ,\ G\ \text{is symmetric}\right\} \qquad (19)$$

where $\varepsilon^*\in[0,1)$ is the contamination degree, G is an arbitrary symmetric distribution and N_D is a normal distribution

$$N_D(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt \qquad (20)$$

The contamination model (19) for probability densities has a form

$$P_\varepsilon=\left\{p:p=\left(1-\varepsilon^*\right)N_p\left(0,\sigma_N^2\right)+\varepsilon^*g(e)\right\} \qquad (21)$$

where

$$N\left(0,\sigma_N^2\right)=p(e)=\frac{1}{\sqrt{2\pi\sigma_N}}\exp\left\{-\frac{e^2}{2\sigma_N^2}\right\} \qquad (22)$$

The robust identification algorithm design is based on robust statistics [19] framework. Important ingredient here is the least favorable probability density for a class of probability densities (21)

$$p^{*}\left(e\left(k\Delta\right)\right)=\begin{cases}\dfrac{1-\varepsilon^{*}}{\sqrt{2\pi}\sigma_{N}}\exp\left\{-\dfrac{e^{2}\left(k\Delta\right)}{2\sigma_{N}^{2}}\right\}, & \left|e\left(k\Delta\right)\right|\le k_{\varepsilon}\\ \dfrac{1-\varepsilon^{*}}{\sqrt{2\pi}\sigma_{N}}\exp\left\{-\dfrac{k_{\varepsilon}}{\sigma_{N}^{2}}\left(\left|e\left(k\Delta\right)\right|-\dfrac{k_{\varepsilon}}{2}\right)\right\}, & \left|e\left(k\Delta\right)\right|>k_{\varepsilon}\end{cases} \tag{23}$$

where the relationship between the contamination degree ε^{*} and parameter k_{ε} of Huber`s function is given with next relations

$$\frac{2\Phi_{N}\left(k_{\varepsilon}\right)}{k_{\varepsilon}}-2\Phi_{N}\left(k_{\varepsilon}\right)=\frac{\varepsilon^{*}}{\varepsilon^{*}-1},\quad \Phi_{N}\left(x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^{2}}{2}}dy \tag{24}$$

Good choice for robust algorithms is $k_{\varepsilon}\in\left[2,4\right]$

The Huber loss function is

$$\Phi\left(\varepsilon\left(k\Delta\right)\right)=-\log p^{*}\left(e\left(k\Delta\right)\right)\Big|_{e(k\Delta)=\varepsilon(k\Delta)} \tag{25}$$

where

$$\varepsilon\left(k\Delta\right)=y\left(k\Delta\right)-\boldsymbol{\varphi}^{T}\left(k\Delta\right)\hat{\boldsymbol{\theta}}\left(k\Delta-\Delta\right) \tag{26}$$

and $\hat{\boldsymbol{\theta}}\left(k\Delta\right)$ is estimate of parameters θ.

From relation (23) and (25) it follows that

$$\Phi\left(\varepsilon\left(k\Delta\right)\right)=\begin{cases}\dfrac{\varepsilon^{2}\left(k\Delta\right)}{2\sigma_{N}^{2}}+\log\dfrac{\sqrt{2\pi}\sigma_{N}}{1-\varepsilon^{*}}, & \left|\varepsilon\left(k\Delta\right)\right|\le k_{\varepsilon}\\ \dfrac{k_{\varepsilon}}{\sigma_{N}^{2}}\left(\left|\varepsilon\left(k\Delta\right)\right|-\dfrac{k_{\varepsilon}}{2}\right)+\log\dfrac{\sqrt{2\pi}\sigma_{N}}{1-\varepsilon^{*}}, & \left|\varepsilon\left(k\Delta\right)\right|>k_{\varepsilon}\end{cases} \tag{27}$$

Derivation of $\Phi\left(\cdot\right)$ is a Huber`s function

$$\psi\left(\varepsilon\left(k\Delta\right)\right)=\begin{cases}\varepsilon\left(k\Delta\right), & \left|\varepsilon\left(k\Delta\right)\right|\le k_{\varepsilon}\\ k_{\varepsilon}, & \left|\varepsilon\left(k\Delta\right)\right|>k_{\varepsilon}\end{cases} \tag{28}$$

From relation (28) it follows that Huber function is not differentiable in the two points $\left(\left(k_{\varepsilon}\right)\text{ and }\left(-k_{\varepsilon}\right)\right)$. It means that Huber loss function (25) is only first – order differentiable and it follows that is not applicable to second order method (such is a Newton – Raphson algorithm, which is considered in this paper). Therefore we take a smooth version of the Huber loss function (pseudo Huber's loss function) which has derivatives of all degrees [21].

In our case pseudo - Huber loss function has a form

$$\Phi_{p}\left(\varepsilon\left(k\Delta\right)\right)=k_{\varepsilon}\left(\sqrt{\left(k_{\varepsilon}\right)^{2}+\varepsilon^{2}\left(k\Delta\right)}-k_{\varepsilon}\right)+\log\frac{\sqrt{2\pi}\sigma}{1-\varepsilon^{*}} \tag{29}$$

The functions $\Phi\left(\cdot\right)$ and $\Phi_{p}\left(\cdot\right)$ are close [20-21]. The derivatives of loss function $\Phi_{p}\left(\cdot\right)$ are

$$\psi_{p}\left(\varepsilon\left(k\Delta\right)\right)=\Phi'_{p}\left(\varepsilon\left(k\Delta\right)\right)=\frac{k_{\varepsilon}\varepsilon\left(k\Delta\right)}{\sqrt{\left(k_{\varepsilon}\right)^{2}+\varepsilon^{2}\left(k\Delta\right)}} \tag{30}$$

$$\psi'_{p}\left(\varepsilon\left(k\Delta\right)\right)=\Phi''_{p}\left(\varepsilon\left(k\Delta\right)\right)=\frac{\left(k_{\varepsilon}\right)^{3}}{\left(\sqrt{\left(\left(k_{\varepsilon}\right)^{2}+\varepsilon^{2}\left(k\Delta\right)\right)}\right)^{3/2}} \tag{31}$$

The derivatives of $\psi_p(\cdot)$ (that is pseudo - Huber function) and $\psi'_p(\cdot)$ are bounded and Lipschitz continuous [21].

According with above discussion and philosophy of Algorithm III from section 2 of the paper we can define next two functionals

$$J(\boldsymbol{\theta}) = E\left\{\sum_{i=1}^{k} \Phi\left(\varepsilon(i\Delta)\right)\right\} \qquad (32)$$

$$J^{p}(\boldsymbol{\theta}) = E\left\{\sum_{i=1}^{k} \Phi_{p}\left(\varepsilon(i\Delta)\right)\right\} \qquad (33)$$

The corresponding empirical functionals are

$$J_k(\boldsymbol{\theta}) = \frac{1}{k}\sum_{j=1}^{k}\sum_{i=1}^{j} \Phi\left(\varepsilon(i\Delta)\right) \qquad (34)$$

$$J_k^{p}(\boldsymbol{\theta}) = \frac{1}{k}\sum_{j=1}^{k}\sum_{i=1}^{j} \Phi_{p}\left(\varepsilon(i\Delta)\right) \qquad (35)$$

We now introduce modified Newton – Raphson method in next form

$$\hat{\boldsymbol{\theta}}(k\Delta) = \hat{\boldsymbol{\theta}}(k\Delta - \Delta) - \left[\nabla_{\theta}^{2} J_k^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right)\right]^{-1}\left[\nabla_{\theta} J_k\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right)\right] = \hat{\boldsymbol{\theta}}(k\Delta - \Delta) - \left[k\nabla_{\theta}^{2} J_k^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right)\right]^{-1}\left[k\nabla_{\theta} J_k\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right)\right] \qquad (36)$$

From relation (34) it follows that

$$J_k(\boldsymbol{\theta}) = \frac{1}{k}\left((k-1)\frac{1}{k-1}\right)\sum_{j=1}^{k-1}\sum_{i=1}^{j} \Phi\left(\varepsilon(i\Delta)\right) + \frac{1}{k}\sum_{i=1}^{k} \Phi\left(\varepsilon(i\Delta)\right) = \frac{k-1}{k} J_{k-1}(\boldsymbol{\theta}) + \frac{1}{k}\sum_{i=1}^{k} \Phi\left(\varepsilon(i\Delta)\right) \qquad (37)$$

From last relation we have

$$kJ_k(\boldsymbol{\theta}) = (k-1) J_{k-1}(\boldsymbol{\theta}) + \sum_{i=1}^{k} \Phi\left(\varepsilon(i\Delta)\right) \qquad (38)$$

By differentiation of relation (38) and take $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(k\Delta - \Delta)$ one can get

$$k\nabla_{\theta} J_k\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) = (k-1)\nabla_{\theta} J_{k-1}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) - \sum_{i=1}^{k} \boldsymbol{\varphi}(i\Delta)\,\psi\left(\varepsilon(i\Delta)\right) \qquad (39)$$

Let us supposed that $\hat{\boldsymbol{\theta}}(k\Delta - \Delta)$ is optimal at $(k\Delta - \Delta)$, *i.e*

$$\nabla_{\theta} J_{k-1}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) = 0 \qquad (40)$$

From (36), (39) and (40) we have

$$\hat{\boldsymbol{\theta}}(k\Delta) = \hat{\boldsymbol{\theta}}(k\Delta - \Delta) + \left[k\nabla_{\theta}^{2} J_k^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right)\right]^{-1}\sum_{i=0}^{k} \boldsymbol{\varphi}(i\Delta)\,\psi\left(\varepsilon(i\Delta)\right) \qquad (41)$$

As for functional (34) one can get for functional (35)

$$k\nabla_{\theta} J_k^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) = (k-1)\,k\nabla_{\theta} J_{k-1}^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) - \sum_{i=0}^{k} \boldsymbol{\varphi}(i\Delta)\,\psi_{p}\left(\varepsilon(i\Delta)\right) \qquad (42)$$

From last relation we have

$$k\nabla_{\theta}^{2} J_k^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) = (k-1)\,k\nabla_{\theta}^{2} J_{k-1}^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) + \sum_{i=0}^{k} \psi'_{p}\left(\varepsilon(i\Delta)\right)\boldsymbol{\varphi}(i\Delta)\,\boldsymbol{\varphi}^{T}(i\Delta) \qquad (43)$$

Let us suppose that $\hat{\boldsymbol{\theta}}(k\Delta)$ lies in neighborhood of $\hat{\boldsymbol{\theta}}(k\Delta - \Delta)$.

From that follows

$$\nabla^{2} J_k^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta)\right) \cong \nabla^{2} J_k^{p}\left(\hat{\boldsymbol{\theta}}(k\Delta - \Delta)\right) \qquad (44)$$

By using relations (43) and (44) one can get

$$\nabla^2 J_k^p\left(\hat{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)\right)=\left(k-1\right)\nabla_\theta^2 J_{k-1}^p\left(\hat{\boldsymbol{\theta}}\left(k\Delta-2\Delta\right)\right)+\sum_{i=0}^{k}\psi'_p\left(\varepsilon\left(i\Delta\right)\right)\boldsymbol{\varphi}\left(i\Delta\right)\boldsymbol{\varphi}^T\left(i\Delta\right) \quad (45)$$

Let us introduce

$$\boldsymbol{P}\left(k\Delta\right)=\left[k\nabla_\theta^2 J_k^p\left(\hat{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)\right)\right]^{-1} \quad (46)$$

From last two relations it follows that

$$\boldsymbol{P}^{-1}\left(k\Delta\right)=\boldsymbol{P}^{-1}\left(k\Delta-\Delta\right)+\sum_{i=0}^{k}\psi'_p\left(\varepsilon\left(i\Delta\right)\right)\boldsymbol{\varphi}\left(i\Delta\right)\boldsymbol{\varphi}^T\left(i\Delta\right) \quad (47)$$

By using relations (41) and (47) it follows that

$$\hat{\boldsymbol{\theta}}\left(k\Delta\right)=\hat{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)+P\left(k\Delta\right)\sum_{i=0}^{k}\boldsymbol{\varphi}\left(i\Delta\right)\psi\left(\varepsilon\left(i\Delta\right)\right) \quad (48)$$

Arithmetic averages of the resulting iterates are

$$\bar{\boldsymbol{\theta}}\left(k\Delta\right)=\frac{1}{k}\sum_{i=1}^{k}\hat{\boldsymbol{\theta}}\left(i\Delta\right) \quad (49)$$

By using last relation we have

$$\bar{\boldsymbol{\theta}}\left(k\Delta\right)=\frac{1}{k}\left(\sum_{i=1}^{k-1}\hat{\boldsymbol{\theta}}\left(i\Delta\right)+\hat{\boldsymbol{\theta}}\left(k\Delta\right)\right)=\frac{k-1}{\left(k-1\right)k}\sum_{i=1}^{k-1}\hat{\boldsymbol{\theta}}\left(i\Delta\right)+\frac{1}{k}\hat{\boldsymbol{\theta}}\left(k\Delta\right)=$$

$$\frac{k-1}{k}\bar{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)+\frac{1}{k}\hat{\boldsymbol{\theta}}\left(k\Delta\right)=\bar{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)-\frac{1}{k}\bar{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)+\frac{1}{k}\hat{\boldsymbol{\theta}}\left(k\Delta\right) \quad (50)$$

We will make modification of gain of stochastic procedure (48)

a) Convert the matrix gain to scalar gain

$$r\left(k\Delta\right)=tr\boldsymbol{P}^{-1}\left(k\Delta\right)$$

b) Further modification of scalar gain according with Algorithm I – Algorithm III

$$\frac{1}{r\left(k\Delta\right)^\gamma}\ ,\quad \frac{1}{2}<\gamma<1$$

By using relations (28), (31), (48), (50 – (52) we have next accelerated algorithm

Algorithm – 1

(Acceleration is based on averaging in both iterates and observations)

- $$\hat{\boldsymbol{\theta}}\left(k\Delta\right)=\hat{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)+\frac{1}{r\left(k\Delta\right)^\gamma}\sum_{i=0}^{k}\boldsymbol{\varphi}\left(i\Delta\right)\psi\left(\varepsilon\left(i\Delta\right)\right)\ ,\quad \frac{1}{2}<\gamma<1 \quad (53)$$
- $$r\left(k\Delta\right)=r\left(k\Delta-\Delta\right)+\sum_{i=0}^{k}\psi'_p\left(\varepsilon\left(i\Delta\right)\right)\left\|\boldsymbol{\varphi}\left(i\Delta\right)\right\|^2 \quad (54)$$
- $$\varepsilon\left(k\Delta\right)=y\left(k\Delta\right)-\boldsymbol{\varphi}^T\left(k\Delta\right)\hat{\boldsymbol{\theta}}\left(k\Delta-\Delta\right) \quad (55)$$
- $$\psi\left(\varepsilon\left(k\Delta\right)\right)=\begin{cases}\varepsilon\left(k\Delta\right)\ , & \left|\varepsilon\left(k\Delta\right)\right|\le k_\varepsilon\\ k_\varepsilon\ , & \left|\varepsilon\left(k\Delta\right)\right|>k_\varepsilon\end{cases}\ ,\quad k_\varepsilon\in\left[2,4\right] \quad (56)$$
- $$\psi'_p\left(\varepsilon\left(k\Delta\right)\right)=\frac{\left(k_\varepsilon\right)^3}{\left(\sqrt{\left(\left(k_\varepsilon\right)^2+\varepsilon^2\left(k\Delta\right)\right)}\right)^{3/2}} \quad (57)$$
- $$\bar{\boldsymbol{\theta}}\left(k\Delta\right)=\bar{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)-\frac{1}{k}\bar{\boldsymbol{\theta}}\left(k\Delta-\Delta\right)+\frac{1}{k}\hat{\boldsymbol{\theta}}\left(k\Delta\right) \quad (58)$$
- Initial conditions

$$\hat{\boldsymbol{\theta}}\left(0\right)=\boldsymbol{0}\ ,\ \bar{\boldsymbol{\theta}}\left(0\right)=\boldsymbol{0}\ ,\ r\left(0\right)=1 \quad (59)$$

■ ■ ■

The problem with the identification of FIR model is that order of model is unknown. For model order estimation we propose robust version of BIC criterion. The algorithm for that criterion has a form

Algorithm – 2

(Robust order determination of FIR model)

- Dimension of parameter

$$p = \dim \boldsymbol{\theta} \quad (60)$$

- N – number of observations
- $$BIC(p,N) = \log\left(\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{j}\Phi\left(\varepsilon(i,p)\right)\right) + \frac{p\log N}{N} \quad (61)$$
- $$\varepsilon(i,p) = y(i) - \boldsymbol{\varphi}^{T}(i,p)\,\bar{\boldsymbol{\theta}}(N,p) \quad (62)$$

$$\Phi\left(\varepsilon(k\Delta)\right) = \begin{cases} \dfrac{\varepsilon^{2}(k\Delta)}{2\sigma_{N}^{2}} + \log\dfrac{\sqrt{2\pi\sigma_{N}}}{1-\varepsilon^{*}}, & \left|\varepsilon(k\Delta)\right| \le k_{\varepsilon} \\ \dfrac{k_{\varepsilon}}{\sigma_{N}^{2}}\left(\left|\varepsilon(k\Delta)\right| - \dfrac{k_{\varepsilon}}{2}\right) + \log\dfrac{\sqrt{2\pi\sigma_{N}}}{1-\varepsilon^{*}}, & \left|\varepsilon(k\Delta)\right| > k_{\varepsilon} \end{cases} \quad (63)$$

- Order of the model

$$\hat{p}(N) = \arg\min_{p} B(p,N) \quad (64)$$

■ ■ ■

Practical identification of FIR model (parameter and order) is based on simultaneous application of both **Algorithm -1** and **Algorithm -2** for different values of p. The minimum of criterion BIC (p,N), for all p, determines the order of the model. For p for which criterion BIC(p,N) has minimal value Algorithm -1 gives the value of parameter θ of FIR model.

Second stage (identification of fast OE model)

In the previous stage it is identified FIR model (18). Using that model the prediction of noise free outputs is

$$y_{a}(k\Delta + j) = \boldsymbol{\varphi}^{T}(k\Delta + j)\,\hat{\boldsymbol{\theta}}(k\Delta) \quad , \quad j = 0,1,2,\ldots,\Delta - 1 \quad (65)$$

According with philosophy of output-error method the values $y_{a}(.)$ can be considered as outputs of auxiliary model.

Let us form the next vectors

$$\boldsymbol{\theta}_{s}^{T} = \left[a_{1}, a_{2}, \ldots a_{n}, b_{1}, b_{2}, \ldots b_{m},\right] \quad (66)$$

$$\boldsymbol{\varphi}_{0}^{T}(k) = \left[-y_{a}(k-1), \ldots, -y(k-n), u(k-1), \ldots, u(k-m)\right] \quad (67)$$

The vector form of model is

$$y(k) = \boldsymbol{\varphi}_{0}(k)\,\boldsymbol{\theta}_{s} + e(k) \quad (68)$$

Identification of parameter $\boldsymbol{\theta}_{s}$ we perform by application of **Algorithm – 1**.For above considerations it follows that the identification of fast rate model consists of two steps: (i) identification of FIR model in dual – rate setting by using accelerated recursive algorithm (Algorithm – 1) and BIC information criterion (Algorithm – 2); (ii) based on output error method philosophy FIR model is used for calculation of model output which, together with input signals, form information vector $\boldsymbol{\varphi}_{0}(\cdot)$. Further is applied Algorithm – 1 for identification of vector parameter $\boldsymbol{\theta}_{s}$.

IV. SIMULATIONS

Let us suppose that fast discrete – time model has a next form.

$$y(k) = \frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k) + e(k)$$

$$A\left(q^{-1}\right) = 1 - 1.5q^{-1} + 0.7q^{-2}$$

$$B\left(q^{-1}\right) = q^{-1} + 0.5q^{-2}$$

The input $\{u(k)\}$ is generated with uniform distribution on $[-1,1]$. It is supposed that stochastic disturbance has a non-Gaussian distribution

$$e \sim \left(1-\varepsilon^{*}\right) N\left(0.\sigma_N^2\right) + \varepsilon^{*} N\left(0.\sigma_2^2\right)$$

where $N\left(0,\sigma^2\right)$ is Gaussian distribution with zero mean and variance σ^2. In all simulation it is supposed that

$$\sigma_N^2 = 1 \;,\; \sigma_N^2 = 100 \;,\; N = 3000 \;,\; \Delta = 2 \;,\; k_\varepsilon = 3$$

The form of estimation error is

$$E_k = \log\left\|\bar{\theta}_s(k) - \theta_s\right\|^2$$

We will consider next types of errors
SA – standard stochastic approximation algorithm (relation (2))
ACC – accelerated stochastic approximation algorithm (Algorithm-1)
First we will determine parameter and order of FIR model. Results of simulations are

p	$BIC(p,N)$
6	1.4988
7	1.4785
8	1.3386
9	1.4571
10	1.4755

We see that order of FIR model is $p = 8$.
The parameter of that FIR model is

$$\boldsymbol{\theta}^T = [1.10, 1.73, 2.40, 1.95, 1.47; 0.79; 0.03; -0.24]$$

Further we consider estimation of parameters $a_i(1,2,\ldots,n)$ and $b_j(1,2,\ldots,m)$

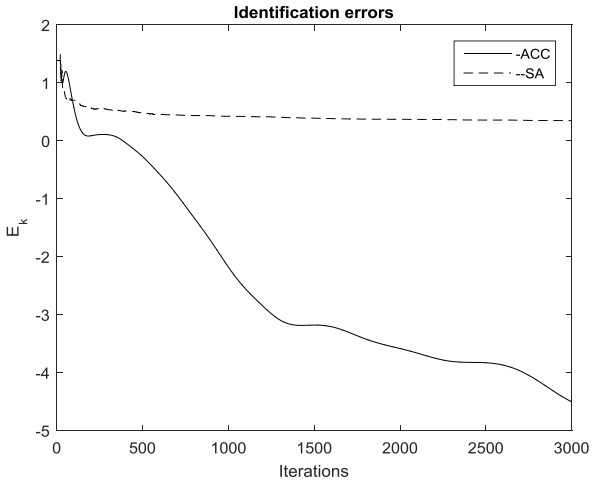

Fig. 1 Comparison of SA and ACC, $\varepsilon^{*} = 0.2 \;,\; \gamma = 0.75$

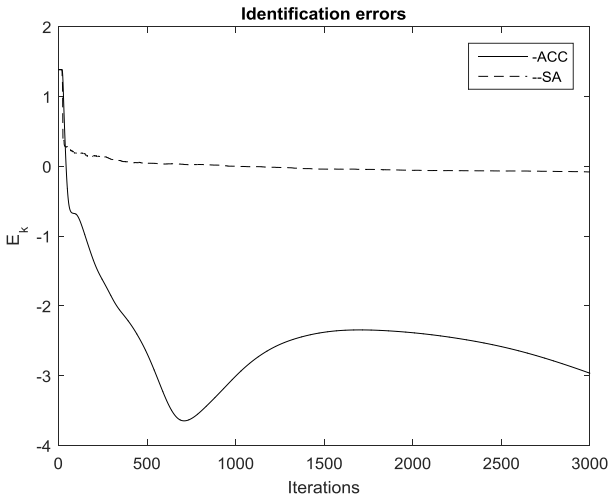

Fig. 2 Comparison of SA and ACC , $\varepsilon^* = 0.1$, $\gamma = 0.75$

Next two figures consider influence of parameter γ estimation errors

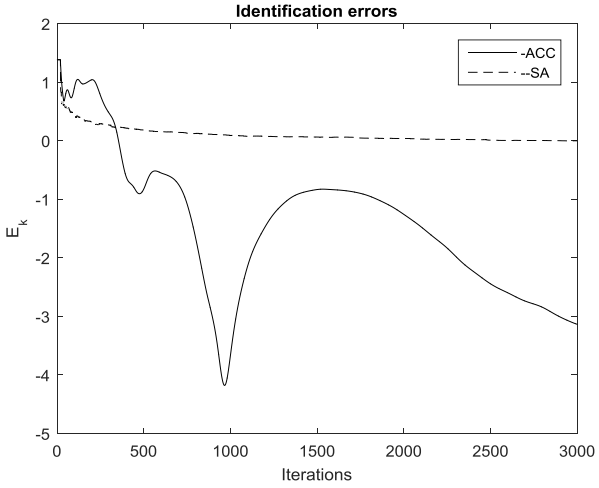

Fig. 3 Comparison of SA and ACC, $\varepsilon^* = 0.2$, $\gamma = 0.6$

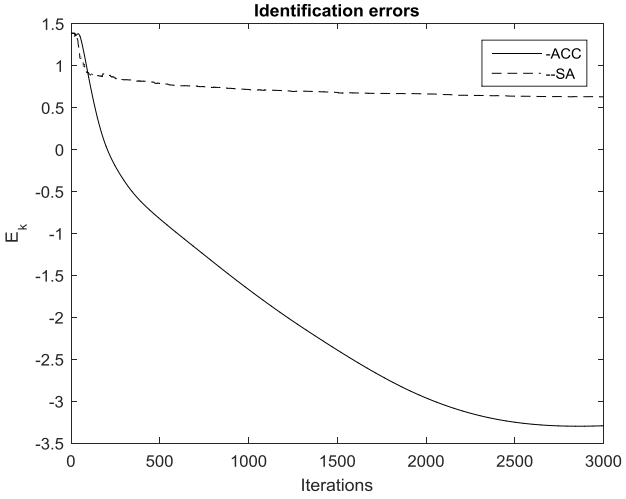

Fig. 4 Comparison of SA and ACC , $\varepsilon^* = 0.2$, $\gamma = 0.85$

We can conclude from simulations that parameter $\gamma \in [0.7 = 0.85]$. The best result in our example is given for γ =0.7

V. CONCLUSIONS

IN this paper are presented robust accelerated stochastic approximation algorithm and robust Bayesian information criterion. The robust identification algorithm is based on modified Newton – Raphson algorithm. The robust Bayesian information criterion is derived by using robust statistics. Based on two above algorithm it is proposed two stage procedure which determines the fast rate model of the dual – rate system. The simulation study shows effectiveness of proposed method.

REFERENCES

[1]. Kranc, G.M.:Input – output analysis of multirate feedback systems. IRE Transactions on Automtic Contro **3**(1), 21 – 28 (1957)
[2]. Ragazzini, J.R., Franklin, G.F.: Sampled – Data Control Systems. McGraw – Hill, New York (1958)
[3]. Shah, L.D. Chen, T., Qi, K.Z.: Application of dual – rate modelling to CCR octane quality inferential control. IEEE Transactions on Control Systems Technology **11**, 43 – 51(2003).doi:101109/TCST.2002.866433.
[4]. Luchini, E., Schirrer, A., S. Jakubek, S., Kozek, M.: Model predictive multirate control for mixed-integer optimization of redundant refrigeration circuits. Journal of Process Control **76**, 112 – 121 (2019). doi:10.1016/j.jproccont.2018.12.015
[5]. Fu,Y., Du, Q., Zhou, X., Fu, J., Chai, T.: Dual – rate adaptive decoupling controller and its application to a dual – tank water system. IEEE Transactions on Control Systems Technology **28**, 2515 – 2522 (2020). doi:10.1109/TCST.2019.2930044.
[6]. Chen, W., Qu,L.: Stabilization of networked control systems with multirate sampling. Automatica **49**, 1528 – 1537 (2013). doi:10.1016/j.automatica.2013.02.010.
[7]. Chen, J., Liu Y, Xu, Y.: A new filter – based stochastic gradient algorithm for dual – rate ARX models. Adaptive Control and Signal Processing **32**, 1557 – 1568 (2018).doi:10.1002/acs.2930
[8]. Li, W., Han, Z., Shan, S.: Subspace identification for FDI in systems with non – uniformly sampled multirate data. Automatica **42**, 610 – 627 (2006). doi:10.1016/j.automatica.2005.11.010.
[9]. Ding, F., Liu, X., Shi, Y.:Convergence analysis of estimation algorithms of dual – rate stochastic systems. Applied Mathematics and Computation **176**, 245 – 261 (3006). doi:1016/jamc.2005.09.048.
[10]. Ljung, L., Soderstrom, T.: Theory and Practice of Recursive Identification. MIT Press, Cambridge Massachusetts (1983)
[11]. Ding, F., Chen, T.: Identification of dual – rate systems based on finite impulse response models. Adaptive Control and Signal Processing **18**, 589 – 598 (2004). doi:10.1002/acs.820
[12]. Ding, F., Chen, T.: Parameter estimation of dual – rate stochastic systems by using an output-error method. IEEE Transactions on Automatic Control **50**, 1430 – 1441 (2005).doi:10.1109/TAC.2005.854654.
[13]. Robbins, H. Monro, S.: A stochastic approximation method. Ann. Math. Statist. **22**, 400-407 (1951)
[14]. Polyak, B.T.: New stochastic approximation type procedures. Automat. Remote Control 57, 98-107 (1990)
[15]. Kushner, H.J., Yin, G.G.: Stochastic Approximation and Recursive Algorithms and Applications. Springer, Berlin (2003)
[16]. Yin, G.: Adaptive filtering with averaging. In: Goodwin GC, Astrom K, Kumar PR (Ed).: Adaptive Control Filtering and Signal Processing. pp: 375 – 396 Springer, Berlin
[17]. Schwarz, G.: Estimating the dimension of a model. The Annals of Statistics **6**, 461 – 464 (1978)
[18]. Chen, B., Zhu, Y., Hu, J. Principe, J.G.: System Parameter Identification. Information Criteria and Algorithms. Elsevier, London (2013)
[19]. Huber, P., Ronchetti,E.: Robust Statistics. Wiley, New York (2009)
[20]. Stefanski, L., Boss, D.: The calculus of M – estimation. The American Statistician **56**, 29-38 (2009).doi: 10.1198/000313002753631330.
[21]. Hotly, R.: Multiple View Geometry in Computer Visions. Cambridge University Press, Cambridge (2004)
[22]. Kesten, H.:Accelerated stochastic approximation. Ann. Math. Statist. **29**, 41 – 59 (1958)
[23]. Yin, G., Yin, K.: Asymptotically optimal rate convergence of smoothed stochastic recursive algorithms. Stochastic and Stochastic Reports **47**, 21 – 46 (1994)
[24]. Kushner, H., Yang, J.: Stochastic approximation with averaging and feedback:rapidly convergent "on – line" algorithms. IEEE Trans. Automatic Control **40**, 24 – 34 (1995)
[25]. Filipovic, V.: A global convergent outlier robust adaptive prediction for MIMO Hammerstein model. Int. J. Robust and Nonlinear Control **27**, 3350 – 3371 (2017). doi;10.1002//rnc.3705
[26]. Filipovic, V.: Consistency of the robust recursive Hammerstein model identification algorithm. J. Franklin Inst. **352**, 1932 – 1943 (2014). doi: 101016//j.jfranklin.201502.005