Vojislav Filipovic. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 12, Issue 2, (Series-I) February 2022, pp. 47-66

RESEARCH ARTICLE OPEN ACCESS

Robust identification of multivariable Hammerstein models based on modified Newton – Raphson method

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

ABSTRACT
The paper considers the robust recursive stochastic algorithm for identification of multivariable Hammerstein model with a general static nonlinear block in polynomial form and a linear block described with multivariable ARMAX (Autoregressive Moving Average with Exogenous Variables) model. It is assumed that there is a priori information about the distribution class to which a real disturbance belongs. Such description of disturbance allows the presence of outliers in observations. Design of recursive algorithm, in this paper, based on robust statistics and Newton – Raphson method. The Newton – Raphson method requires that the loss function should be twice differentiable. Huber loss function, based on robust statistics, has just first derivative. The problem can be overcome with the pseudo – Huber loss function which has the derivatives of arbitrary orders and which behaves similarly as Huber loss function. In this paper the Huber loss function is used for first derivative of functional while pseudo - Huber loss function is used for second derivative of functional. That is essence of modification of Newton – Raphson method. The main contributions of the paper are: (i) Design a new algorithm for identification of multivariable nonlinear systems; (ii) The convergence analysis.

Keywords: Hammerstein model˙ outliers˙ Huber theory, pseudo - Huber theory ˙ convergence Analysis

---

Date of Submission: 05-02-2022 Date of Acceptance: 18-02-2022

I.

II. INTRODUCTION

Multivariable systems represent an important class of practical systems [1]. The examples are thermal power plants and distillation columns. Designing of regulators for such processes requires their mathematical models. From the investigations point of view identification theory is vibrant [2-4]. A great deal of attention is dedicated to identification of multivariable stochastic systems [5,6]. The key role of process modelling has a nature of disturbance. Most frequently it is supposed that disturbance has Gaussian distribution. Practical research showed that this assumption not justified [7]. Namely, in population of observations there are rare large observations (outliers). Owing that fact it is imperative to design recursive identification algorithm which has minor sensitivity to presence of outliers. The main tool for design of such algorithms is a robust statistics [8-9]. Using this theory it is possible to get robust recursivealgorithms.

Review of recent result in robustindustrial process identification is given in reference [10] where the different probabilistic methods are used for outlier modelling. The concept of influence function [11] also can be used in robust estimation. This concept is considered in references [12, 13]

Robust estimation is relevant for different fields: (i) outliers detection based on data mining perspective [14, 15]; (ii) machine learning [16, 17]; (iii) signal processing [18]; (iv) principal component analysis [19].

Robust identification and prediction of multivariable systems are considered in [20 – 22].

In this papera new robust algorithm is derived. Namely, the robust identification is based on Huber loss function depending on the most unfavourable density of probability of disturbance. This function has only derivatives of first order. The Huber function (derivative of Huber loss function) is not differentiable in two points $(+k_\varepsilon)$ and $(-k_\varepsilon)$ where k_ε is parameter of the Huber function. From that fact it follows that Huber loss function is not applicable to second order methods. (for example,the Newton – Raphson algorithm, which is considered in this paper). In reference [23] the problem has been overcome by upper bound for Fisher information. In this paper it is used pseudo -

Huber loss function [24 – 26] which has derivatives of arbitrary order.
In this paper is proposed modified Newton – Raphson algorithm in which Huber loss function is used for the first derivative of functional while for second derivative of functional the pseudo - Huber loss function is used. The algorithms has gain matrix which explicitly depends on the second derivative of pseudo - Huber loss function.

The convergence analysis of given algorithm is based on martingale theory [27] and theory of passive operators [28].

The main contributions of the paper are:

(i) Design of new robust recursive algorithm, for identification of multivariable Hammerstein model, based on modified Newton – Raphson method.

(ii) The convergence analysis of algorithm.

The paper is organized as follows. Section 2 describes the multivariable Hammerstein model of the system. In section 3 is considered derivation of algorithms based on Huber loss function and pseudo – Huber loss function. Section 4 considers convergence analysis of robust algorithm and behaviour of algorithmis considered in section 5 (Simulation study). The concluding remarks are given in the last section.

III. MULTIVARIABLE NARMAX MODEL

Let us suppose that the system is described by the nonlinear multivariable ARMAX model with r – dimensional input and p – dimensional output.

$$\boldsymbol{A}\left(q^{-1}\right)\boldsymbol{y}_k = \boldsymbol{B}\left(q^{-1}\right)\boldsymbol{f}\left(\boldsymbol{u}_k\right) + \boldsymbol{C}\left(q^{-1}\right)\boldsymbol{w}_k \quad (1)$$

where $\boldsymbol{A}\left(q^{-1}\right)$ and $\boldsymbol{B}\left(q^{-1}\right)$ and $\boldsymbol{C}\left(q^{-1}\right)$ are matrix polynomials and q^{-1} denotes the shift – back operator $\left(q^{-1}x_n = x_{n-1}\right)$. Order of polynomials $\boldsymbol{A}\left(q^{-1}\right)$, $\boldsymbol{B}\left(q^{-1}\right)$ and $\boldsymbol{C}\left(q^{-1}\right)$ are

$$\boldsymbol{A}\left(q^{-1}\right) = \boldsymbol{I} + \boldsymbol{A}_1 q^{-1} + \ldots + \boldsymbol{A}_n q^{-n}$$

$$\boldsymbol{B}\left(q^{-1}\right) = \boldsymbol{B}_1 q^{-1} + \ldots + \boldsymbol{B}_m q^{-m}$$

$$\boldsymbol{C}\left(q^{-1}\right) = \boldsymbol{I} + \boldsymbol{C}_1 q^{-1} + \ldots + \boldsymbol{C}_l q^{-l} \quad (2)$$

where $\boldsymbol{A}_i\left(i = 1,2,\ldots,n\right)$ are r x r matrices, $\boldsymbol{B}_i\left(i = 1,2,\ldots,m\right)$ are r x p matrices and $\boldsymbol{C}_i\left(i = 1,2,\ldots,l\right)$ are r x r matrices. The stochastic disturbance $\left\{\boldsymbol{w}_k\right\}$ is a martingale difference in relation to thenondecreasing family of σ - algebras $\left\{F_n\right\}$.

Function $\boldsymbol{f}\left(\boldsymbol{u}_k\right)$ in model (1) is a nonlinear vector function and is introduced in [23].

$$\boldsymbol{f}\left(\boldsymbol{u}_k\right) = \left[f_1\left(u_k^1\right), f_2\left(u_k^2\right), \ldots, f_r\left(u_k^r\right)\right]^T, \quad \boldsymbol{f}\left(\boldsymbol{u}_k\right) \in R^r \quad (3)$$

The $f_i\left(u_k^i\right)\ \left(i = 1,2,\ldots,r\right)$ are a nonlinear functions of a known basis ($\gamma_1, \gamma_2, \ldots, \gamma_s$).

$$f_i\left(u_k^i\right) = d_1^i \gamma_1\left(u_k^i\right) + d_2^i \gamma_2\left(u_k^i\right) + \ldots + d_{n_j}^i \gamma_{n_j}\left(u_k^i\right) \quad (4)$$

where $d_j^i \left(i=1,2,...,r;\quad j=1,2,...,n_j\right)$ are unknown parameter.
We will introduce

$$s=\max_j\left\{n_j\right\} \quad (5)$$

From relations (4) and (5) we have

$$\boldsymbol{f}\left(\boldsymbol{u}_k\right)=\begin{bmatrix} f_1\left(u_k^1\right) \\ f_2\left(u_k^2\right) \\ \vdots \\ f_r\left(u_k^r\right)\end{bmatrix}=\begin{bmatrix} d_1^1\gamma_1\left(u_k^1\right)+d_2^1\gamma_2\left(u_k^1\right)+...+d_s^1\gamma_s\left(u_k^1\right) \\ d_1^2\gamma_1\left(u_k^2\right)+d_2^2\gamma_2\left(u_k^2\right)+...+d_s^2\gamma_s\left(u_k^2\right) \\ \vdots \\ d_1^r\gamma_1\left(u_k^r\right)+d_2^r\gamma_2\left(u_k^r\right)+...+d_s^r\gamma_s\left(u_k^r\right)\end{bmatrix} \quad (6)$$

where some matrix elements, according to relation (5), are equal to zero.

$$\boldsymbol{D}_i=\begin{bmatrix} d_i^1 & & 0 \\ & d_i^2 & \\ & \ddots & \\ 0 & & d_i^r\end{bmatrix}\ ,\ \ i=1,2,...,s \quad (7)$$

$$\boldsymbol{\Gamma}_i\left(\boldsymbol{u}_k\right)=\begin{bmatrix}\gamma_i\left(u_k^1\right) \\ \gamma_i\left(u_k^2\right) \\ \vdots \\ \gamma_i\left(u_k^r\right)\end{bmatrix}\ ,\ \ i=1,2,...,s \quad (8)$$

From relations (7) and (8) it follows

$$\boldsymbol{D}_i\boldsymbol{\Gamma}_i\left(\boldsymbol{u}_k\right)=\begin{bmatrix} d_i^1\gamma_i\left(u_k^1\right) \\ d_i^2\gamma_i\left(u_k^2\right) \\ \vdots \\ d_i^r\gamma_i\left(u_k^r\right)\end{bmatrix}\ ,\ \ i=1,2,...,s \quad (9)$$

Using relation (9) one can get

$$\boldsymbol{f}\left(\boldsymbol{u}_k\right)=\sum_{i=1}^{s}\boldsymbol{D}_i\boldsymbol{\Gamma}_i\left(\boldsymbol{u}_k\right) \quad (10)$$

For $\boldsymbol{\Gamma}_i\left(\boldsymbol{u}_k\right)$ we have

$$q^{-d}\boldsymbol{\Gamma}_i\left(\boldsymbol{u}_k\right)=\boldsymbol{\Gamma}_i\left(\boldsymbol{u}_{k-d}\right)\ ,\ \ \mathrm{d}=1,2,...,m \quad (11)$$

From relations (2), (10) and (11) it follows that

$$\begin{aligned}\boldsymbol{B}\left(q^{-1}\right)\boldsymbol{f}\left(\boldsymbol{u}_k\right)=&\left[\boldsymbol{B}_1\boldsymbol{D}_1\boldsymbol{\Gamma}_1\left(\boldsymbol{u}_{k-1}\right)+\boldsymbol{B}_2\boldsymbol{D}_1\boldsymbol{\Gamma}_1\left(\boldsymbol{u}_{k-2}\right)+...+\boldsymbol{B}_m\boldsymbol{D}_1\boldsymbol{\Gamma}_1\left(\boldsymbol{u}_{k-m}\right)\right]+\\ &\left[\boldsymbol{B}_1\boldsymbol{D}_2\boldsymbol{\Gamma}_2\left(\boldsymbol{u}_{k-1}\right)+\boldsymbol{B}_2\boldsymbol{D}_2\boldsymbol{\Gamma}_2\left(\boldsymbol{u}_{k-2}\right)+...+\boldsymbol{B}_m\boldsymbol{D}_2\boldsymbol{\Gamma}_2\left(\boldsymbol{u}_{k-m}\right)\right]+...+\\ &\left[\boldsymbol{B}_1\boldsymbol{D}_s\boldsymbol{\Gamma}_s\left(\boldsymbol{u}_{k-1}\right)+\boldsymbol{B}_2\boldsymbol{D}_s\boldsymbol{\Gamma}_s\left(\boldsymbol{u}_{k-2}\right)+...+\boldsymbol{B}_m\boldsymbol{D}_s\boldsymbol{\Gamma}_s\left(\boldsymbol{u}_{k-m}\right)\right]\end{aligned} \quad (12)$$

Now we will introduce the vector

$$\begin{aligned}\left(\boldsymbol{x}_k^0\right)^T=&\left[-\boldsymbol{y}_{k-1}^T,-\boldsymbol{y}_{k-2}^T,...,-\boldsymbol{y}_{k-n}^T,\Gamma_1^T\left(\boldsymbol{u}_{k-1}\right),\boldsymbol{\Gamma}_1^T\left(\boldsymbol{u}_{k-2}\right),...,\boldsymbol{\Gamma}_1^T\left(\boldsymbol{u}_{k-m}\right),\right.\\ &\left.\boldsymbol{\Gamma}_2^T\left(\boldsymbol{u}_{k-1}\right),...,\boldsymbol{\Gamma}_2^T\left(\boldsymbol{u}_{k-m}\right),...,\boldsymbol{\Gamma}_s^T\left(\boldsymbol{u}_{k-1}\right),...,\boldsymbol{\Gamma}_s^T\left(\boldsymbol{u}_{k-m}\right),\boldsymbol{w}_{k-1}^T,...,\boldsymbol{w}_{k-l}^T\right]\end{aligned} \quad (13)$$

and matrix of parameter

$$\left(\boldsymbol{\theta}^M\right)^T = [\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_n, \boldsymbol{B}_1\boldsymbol{D}_1, \boldsymbol{B}_2\boldsymbol{D}_1, ..., \boldsymbol{B}_m\boldsymbol{D}_1, \boldsymbol{B}_1\boldsymbol{D}_2, \boldsymbol{B}_2\boldsymbol{D}_2, ..., \boldsymbol{B}_m\boldsymbol{D}_2, ..., \boldsymbol{B}_1\boldsymbol{D}_s, \boldsymbol{B}_2\boldsymbol{D}_s, ..., \boldsymbol{B}_m\boldsymbol{D}_s, \boldsymbol{C}_1, ..., \boldsymbol{C}_l] \quad (14)$$

Using (13) and (14) we have

$$\boldsymbol{y}_k = \left(\boldsymbol{\theta}^M\right)^T \boldsymbol{x}_k^0 + \boldsymbol{w}_k \quad (15)$$

Let us introducethe matrix

$$\boldsymbol{\varphi}_k = \begin{bmatrix} \left(\boldsymbol{x}_k^0\right)^T & & 0 \\ & \ddots & \\ 0 & & \left(\boldsymbol{x}_k^0\right)^T \end{bmatrix} = \boldsymbol{I} \otimes \left(\boldsymbol{x}_k^0\right)^T \quad (16)$$

where the symbol $\otimes$ denotes the Kroneckerproduct.

Further we will introduce operator "vec" as the operator that generates a column vector by setting the columns of the matrix one below the other.

From (14) we have

$$\boldsymbol{\theta} = vec\ \boldsymbol{\theta}^M \quad (17)$$

Now model (15) has a form

$$\boldsymbol{y}_k = \boldsymbol{\varphi}_k^0 \boldsymbol{\theta}_k + \boldsymbol{w}_k \quad (18)$$

In the above equation matrix $\boldsymbol{\varphi}_k^0$ depends from the immeasurable quantity $\boldsymbol{w}_i \left(i = k-1, k-2, ..., k-l\right)$. The standard action in identification is to replace w_i with the estimated prediction error $\boldsymbol{e}_k$. Now we have

$$x_k^T = \left[-\boldsymbol{y}_{k-1}^T, -\boldsymbol{y}_{k-2}^T, ..., -\boldsymbol{y}_{k-n}^T, \Gamma_1^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\Gamma}_1^T\left(\boldsymbol{u}_{k-2}\right), ..., \boldsymbol{\Gamma}_1^T\left(\boldsymbol{u}_{k-m}\right), \boldsymbol{\Gamma}_2^T\left(\boldsymbol{u}_{k-1}\right), ..., \boldsymbol{\Gamma}_2^T\left(\boldsymbol{u}_{k-m}\right), ..., \boldsymbol{\Gamma}_s^T\left(\boldsymbol{u}_{k-1}\right), ..., \boldsymbol{\Gamma}_s^T\left(\boldsymbol{u}_{k-m}\right), \boldsymbol{e}_{k-1}^T, ..., \boldsymbol{e}_{k-l}^T\right] \quad (19)$$

Now matrix of observation has a form

$$\boldsymbol{\varphi}_k = \begin{bmatrix} \boldsymbol{x}_k^T & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{x}_k^T \end{bmatrix} \quad (20)$$

and prediction error is

$$\boldsymbol{e}_k = \boldsymbol{y}_k - \boldsymbol{\varphi}_k \boldsymbol{\theta}_{k-1} \quad (21)$$

and will be used for definition of identification criterion.

IV. ROBUST RECURSIVE STOCHASTIC ALGORITHM

In this paper we will suppose that the only distribution class to which the stochastic disturbance $\boldsymbol{w}_k$ belongs is known. The general description of such class is

$$P_\varepsilon = \left\{P : P = \left(1-\varepsilon\right)N + \varepsilon G \ \ ,\ G \text{ is symmetric}\right\} \quad (22)$$

where $\varepsilon \in \left[0,1\right)$ is the contamination degree and $N\left(0,\Sigma\right)$ as a zero-mean Gaussian distribution with a covariance matrix Σ. The distribution (22) can describe the presence of outliers in observations.

We will assume that the components of vector $\boldsymbol{w}_k^T = \left[w_k^1 w_k^2 .... w_k^p\right]$ are independent. Using Huber methodology [8] it is possible to find the least favourable probability density on a class (22)

$$p_i^*\left(w_k^i\right)=\begin{cases}\dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_i}\exp\left\{-\dfrac{\left(w_k^i\right)^2}{2\sigma_i^2}\right\}, & \left|w_k^i\right|\le k_\varepsilon^i\\ \dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_i}\exp\left\{-\dfrac{k_\varepsilon^i}{\sigma_i^2}\left(\left|w_k^i\right|-\dfrac{k_\varepsilon^i}{2}\right)\right\}, & \left|w_k^i\right|> k_\varepsilon^i\end{cases} \quad (23)$$

The relation between the contamination degree ε and the parameter k_ε^i of Huber function is

$$\frac{2\Phi_N\left(k_\varepsilon^i\right)}{k_\varepsilon^i}-2\Phi_N\left(-k_\varepsilon^i\right)=\frac{\varepsilon}{\varepsilon-1}$$

$$\Phi_N\left(x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{x^2}{2}}dy \quad (24)$$

Intensive simulations [29] show that the ε move up to 0.2 and $k_\varepsilon^i\in\left[2,4\right]$. The best performance is accomplished for $k_\varepsilon^i=3$.

Let us define the next functions

$$\Phi^i\left(e_k^i\right)=-\ln p_i^*\left(w_k^i\right)\Big|_{w_k^i=e_k^i}=$$

$$=\begin{cases}\dfrac{\left(e_k^i\right)^2}{2\sigma_i^2}+\ln\dfrac{\sqrt{2\pi}\sigma_i}{1-\varepsilon}, & \left|\varepsilon_k^i\right|\le k_\varepsilon^i\\ \dfrac{k_\varepsilon^i}{\sigma_i^2}\left(\left|e_k^i\right|-\dfrac{k_\varepsilon^i}{2}\right)+\ln\dfrac{\sqrt{2\pi}\sigma_i}{1-\varepsilon}, & \left|\varepsilon_k^i\right|> k_\varepsilon^i\end{cases}$$

$$i=1,2,...,p \quad (25)$$

The component of the vector $\boldsymbol{w}_k$, are independent and the least favorable probability density of stochastic disturbance is

$$p^*\left(\boldsymbol{w}_k\right)=\prod_{i=1}^{p}p_i^*\left(w_k^i\right) \quad (26)$$

By using maximum likelihood methodology it is possible to define Huber loss function.

$$\Phi\left(\boldsymbol{x}\right)=-\log p^*\left(\boldsymbol{x}\right),\ \Phi\left(\cdot\right):R^p\rightarrow R^1 \quad (27)$$

From (21) and (27) it follows that identification criterion is.

$$J\left(\boldsymbol{\theta}\right)=E\left\{\Phi\left(\boldsymbol{e}_k\right)\right\} \quad (28)$$

where $E\left\{\cdot\right\}$ is mathematical expectation operator.

In this paper we consider recursive minimization of criterion (28) by applying the Newton – Raphson algorithm. First we introduce the empirical functional for relation (28).

$$J_k\left(\boldsymbol{\theta}\right)=\frac{1}{k}\sum_{i=1}^{k}\Phi\left(\boldsymbol{e}_i\right) \quad (29)$$

The Newton – Raphson algorithm has a form.

$$\boldsymbol{\theta}_k=\boldsymbol{\theta}_{k-1}-\left[k\nabla_\theta^2J_k\left(\boldsymbol{\theta}_{k-1}\right)\right]^{-1}\left[k\nabla_\theta J_k\left(\boldsymbol{\theta}_{k-1}\right)\right] \quad (30)$$

As in [20] from (29) it follows that

$$k\nabla_\theta J_k\left(\boldsymbol{\theta}\right)=\nabla_\theta\Phi\left(\boldsymbol{e}_k\right) \quad (31)$$

The next result will be of interest.

Result 1: [20]. Let us define $\boldsymbol{x} \in R^m$, the vector function $\boldsymbol{y}(\boldsymbol{x}) \in R^n$ and real function $f(\boldsymbol{y}) \in R^1$. Then

$$\frac{\partial f(\boldsymbol{y}(\boldsymbol{x}))}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}^T(\boldsymbol{x})}{\partial \boldsymbol{x}} \frac{\partial f(\boldsymbol{y})}{\partial \boldsymbol{y}} \ \blacksquare$$

From (25) – (27) it follows

$$\Phi(\boldsymbol{e}_k) = -\log p^*(\boldsymbol{w}_k)_{|w_k=e_k} = -\log p_1^*(w_k^1)_{|w_k^1=e_k^1} - \log p_2^*(w_k^2)_{|w_k^2=e_k^2}$$
$$-\ldots - \log p_p^*(w_k^p)_{|w_k^p=e_k^p} = \Phi^1(e_k^1) + \Phi^2(e_k^2) + \ldots + \Phi^p(e_k^p) \quad (32)$$

From (31) and Result 1 we have

$$\nabla_\theta \Phi(e_k) = \frac{\partial \boldsymbol{e}_k^T}{\partial \boldsymbol{\theta}} \frac{\partial \Phi(\boldsymbol{e}_k)}{\partial \boldsymbol{e}_k} \quad (33)$$

By using relations (21) one can get

$$\frac{\partial \boldsymbol{e}_k^T}{\partial \boldsymbol{\theta}} = -\boldsymbol{\varphi}_k^T \quad (34)$$

On the other hand from relation (32) it follows that

$$\frac{\partial \Phi(\boldsymbol{e}_k)}{\partial \boldsymbol{e}_k} = \frac{\partial \Phi^1(e_k^1)}{\partial e_k^1} + \frac{\partial \Phi^2(e_k^2)}{\partial e_k^2} + \ldots + \frac{\partial \Phi^p(e_k^p)}{\partial e_k^p} =$$
$$= \begin{bmatrix} \psi^1(e_k^1) \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \psi^2(e_k^2) \\ \vdots \\ 0 \end{bmatrix} + \ldots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \psi^p(e_k^p) \end{bmatrix} = \begin{bmatrix} \psi^1(e_k^1) \\ \psi^2(e_k^2) \\ \vdots \\ \psi^p(e_k^p) \end{bmatrix} \quad (35)$$

where

$$\psi^i(e_k^i) = \begin{cases} e_k^i & , \ |e_k^i| \le k_\varepsilon^i \\ k_\varepsilon^i \mathrm{sign} e_k^i & , \ |e_k^i| > k_\varepsilon^i \end{cases}$$
$$i = 1, 2, \ldots, p \quad (36)$$

is Huberfunction.

From (31), (33), (34) and (35) it follows that

$$k \nabla_\theta J_k(\boldsymbol{\theta}_{k-1}) = \boldsymbol{\varphi}_k^T \boldsymbol{\psi}(\boldsymbol{e}_k) \quad (37)$$

where

$$\boldsymbol{\psi}(\boldsymbol{e}_k) = \begin{bmatrix} \psi^1(e_k^1) \\ \psi^2(e_k^2) \\ \vdots \\ \psi^p(e_k^p) \end{bmatrix}$$

From relation (36) it follows that Huber`s function $\psi^i(\cdot)$ is not differentiable in the two points (k_ε^i) and $(-k_\varepsilon^i)$. It means that the function $\Phi(\cdot)$ is only first order differentiable and it is impossible to find second derivative of $J_k(\boldsymbol{\theta})$, i.e. $\nabla_\theta^2 J_k(\boldsymbol{\theta})$ which is need for Newton – Raphson method. Because we consider a smooth version of Huber loss function, the pseudo - Huber loss function which has a derivative of all degree [24 – 26].

In our case the pseudo - Huber`s loss function has a form

$$\Phi_{ph}^{i}\left(e_k^i\right)=k_{\varepsilon}^{i}\left(\sqrt{\left(k_{\varepsilon}^{i}\right)^2+\left(e_k^i\right)^2}-k_{\varepsilon}^{i}\right)+\ln\frac{\sqrt{2\pi}\sigma_i}{1-\varepsilon}$$

$$i=1,2,...,p \tag{38}$$

The pseudo - Huber`s loss function is

$$\Phi_{ph}\left(e_k\right)=\Phi_{ph}^{1}\left(e_k^1\right)+\Phi_{ph}^{2}\left(e_k^2\right)+...+\Phi_{ph}^{p}\left(e_k^p\right) \tag{39}$$

Empirical functional for that case has a form

$$J_k^{ph}\left(\boldsymbol{\theta}\right)=\frac{1}{k}\sum_{i=1}^{k}\Phi_{ph}\left(\boldsymbol{e}_i\right) \tag{40}$$

As in relation (35) we have

$$\frac{\partial\Phi_{ph}\left(\boldsymbol{e}_k\right)}{\partial\boldsymbol{e}_k}=\begin{bmatrix}\psi_{ph}^{1}\left(e_k^1\right)\\ \psi_{ph}^{2}\left(e_k^2\right)\\ \vdots\\ \psi_{ph}^{p}\left(e_k^p\right)\end{bmatrix} \tag{41}$$

Where.

$$\psi_{ph}^{i}=\frac{k_{\varepsilon}^{i}e_k^i}{\sqrt{\left(k_{\varepsilon}^{i}\right)^2+\left(e_k^i\right)^2}}$$

$$i=1,2,...,p \tag{42}$$

From relation (40), similarly as in relation (37) on can get.

$$k\nabla_{\theta}J_k^{ph}\left(\boldsymbol{\theta}_{k-1}\right)=\boldsymbol{\varphi}_k^T\boldsymbol{\psi}_{ph}\left(\boldsymbol{e}_k\right) \tag{43}$$

where

$$\boldsymbol{\psi}_{ph}\left(\boldsymbol{e}_k\right)=\begin{bmatrix}\psi_{ph}^{1}\left(e_k^1\right)\\ \psi_{ph}^{2}\left(e_k^2\right)\\ \vdots\\ \psi_{ph}^{p}\left(e_k^p\right)\end{bmatrix}$$

The second derivative of empirical functional (40), by using relation (43), is

$$\nabla_{\theta}^2J_k^{ph}\left(\boldsymbol{\theta}_{k-1}\right)=\frac{1}{k}\sum_{i=1}^{k}\nabla_{\theta}\left(\boldsymbol{\varphi}_k^T\boldsymbol{\psi}_{ph}\left(\boldsymbol{e}_i\right)\right)=\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\varphi}_k^T\boldsymbol{\psi}'_{ph}\left(\boldsymbol{e}_i\right)\boldsymbol{\varphi}_i \tag{44}$$

For large i the prediction errors e_i are approximately independent and matrix $\boldsymbol{\psi}'_{ph}\left(\boldsymbol{e}_i\right)$ is diagonal matrix. Let us introduce $\boldsymbol{\psi}'_{ph}\left(\boldsymbol{e}_i\right)=\boldsymbol{N}\left(\boldsymbol{e}_i\right)$. From relation (44) it follows that

$$\nabla_{\theta}^2J_k^{ph}\left(\boldsymbol{\theta}_{k-1}\right)=\sum_{i=1}^{k}\boldsymbol{\varphi}_k^T\boldsymbol{N}\left(\mathbf{e}_i\right)\boldsymbol{\varphi}_i \tag{45}$$

Where

$$N(e_i)=\begin{bmatrix}\left(\psi_{ph}^{1}\left(e_i^{1}\right)\right)' & & 0\\ & \ddots & \\ 0 & & \left(\psi_{ph}^{p}\left(e_i^{p}\right)\right)'\end{bmatrix} \quad (46)$$

$$\left(\psi_{ph}^{i}\left(e_k^{i}\right)\right)'=\frac{\left(k_\varepsilon^{i}\right)^3}{\left(\sqrt{\left(k_\varepsilon^{i}\right)^2+\left(e_k^{i}\right)^2}\right)^{3/2}}$$

$$i=1,2,...,p \quad (47)$$

In what follows it will be used modified version of Newton – Raphson algorithm (30)

$$\theta_k=\theta_{k-1}-\left[k\nabla_\theta^2 J_k^{ph}\left(\theta_{k-1}\right)\right]^{-1}\left[k\nabla_\theta J_k\left(\theta_{k-1}\right)\right] \quad (48)$$

Let us introduce

$$P_k=\left[k\nabla_\theta^2 J_k^{ph}\left(\theta_{k-1}\right)\right]^{-1} \quad (49)$$

From relations (30), (37), (48) and (49) it follows that

$$\theta_k=\theta_{k-1}+P_k\varphi_k^T\psi\left(e_k\right) \quad (50)$$

By using relations (44) and (49) one can get

$$P_k^{-1}=\sum_{i=1}^{k}\varphi_k^T N\left(e_i\right)\varphi_i=P_{k-1}^{-1}+\varphi_k^T N\left(e_k\right)\varphi_k \quad (51)$$

Now it will be used matrix inversion lemma [30]

$$\left(A+BC\right)^{-1}=A^{-1}-A^{-1}B\left(I+CA^{-1}B\right)^{-1}CA^{-1} \quad (52)$$

with compatible matrix dimensions of matrices ***A***, ***B*** and ***C***.
From (51) it follows that

$$P_k=\left[P_{k-1}^{-1}+\varphi_k^T N\left(e_k\right)\varphi_k\right]^{-1} \quad (53)$$

Let us denote

$$A=P_k^{-1}\ ,\ B=\varphi_k^T\ ,\ C=N\left(e_k\right)\varphi_k \quad (54)$$

Known facts from matrix theory are

$$\left(A^{-1}\right)^{-1}=P_k^{-1}\ ,\ \left(AB\right)^{-1}=B^{-1}A^{-1} \quad (55)$$

From last facts we have

$$\left(I+N\left(e_k\right)\varphi_k P_{k-1}^T\right)^{-1}N\left(e_k\right)=\left(I+N\left(e_k\right)\varphi_k P_{k-1}\varphi_k^T\right)N\left(e_k^{-1}\right)^{-1}$$
$$=\left[N^{-1}\left(e_k\right)\left(I+N\left(e_k\right)\varphi_k P_{k-1}^T\right)\right]^{-1}=\left(N^{-1}\left(e_k\right)+\varphi_k P_{k-1}\varphi_k^T\right)^{-1} \quad (56)$$

By using relations (51) – (59) we finally have relation for matrix gain

$$P_k=P_{k-1}-P_{k-1}\varphi_k^T\left[\varphi_k P_{k-1}\varphi_k^T+N^{-1}\left(e_k\right)\right]^{-1}\varphi_k P_k \quad (57)$$

Relations (50) and (57) represent modified Newton – Raphson algorithm for parameter estimation.

Remark 1. In relation (22) it is given general form of class of distributions. Very often, in practice, it is used Tukey`s class of distributions [31]

$$P=\left(1-\varepsilon\right)N\left(0,\Sigma_1\right)+\varepsilon N\left(0-\Sigma_2\right)\ ,\ \varepsilon\in\left[0,1\right)$$

where $N\left(0,\Sigma_i\right)\ i=1,2$ is the normal distribution with zero mean and the covariance matrix $\Sigma_i\left(i=1,2\right)$ and $\Sigma_1 \ll \Sigma_2$.

Remark 2. We will now show another form of the algorithm (50) and (57). Second derivative of functional $J^{ph}(\boldsymbol{\theta})$ is

$$\nabla_{\theta}^{2} J^{ph}(\boldsymbol{\theta}) = E\left\{\boldsymbol{\varphi}_k^T \boldsymbol{N}(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\right\} = \left\{trE\left(\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^T \boldsymbol{N}(\boldsymbol{e}_k)\right)\right\} \quad (58)$$

Near the minimum of functional $J^{ph}(\boldsymbol{\theta})$ is $\boldsymbol{e}_k \cong \boldsymbol{w}_k$ and relation (58) becomes

$$\nabla_{\theta}^{2} J^{ph}(\boldsymbol{\theta}) = trE\left\{\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^T \boldsymbol{N}(\boldsymbol{w}_k)\right\} = tr\left\{E\left(\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^T\right)E\left(\boldsymbol{N}(\boldsymbol{w}_k)\right)\right\} \quad (59)$$

The $E\left(\boldsymbol{N}(\boldsymbol{w}_k)\right)$ cannot be calculated analytically and, owing that, we introduce approximation for second derivative of pseudo – Huber loss function. Let us introduce

$$\psi_a^i = \begin{cases} 1 \quad , \quad \left|w_k^i\right| \le k_\varepsilon^i \\ 0 \quad , \quad otherwise \end{cases}$$

$$i = 1, 2, ..., p \quad (60)$$

and

$$\boldsymbol{N}_a(\boldsymbol{w}_k) = \begin{bmatrix} \psi_a^1(w_k) & & 0 \\ & \ddots & \\ 0 & & \psi_a^p(w_k) \end{bmatrix} \quad (61)$$

Further is

$$E\left\{\boldsymbol{N}(\boldsymbol{w}_k)\right\} \cong E\left\{\boldsymbol{N}_a(\boldsymbol{w}_k)\right\} = \boldsymbol{M} \quad (62)$$

From last relation it follows that

$$m_i = E\left\{\psi_a^i(w_k)\right\} = \int_{-\infty}^{\infty} \psi_a^i\left(w_k^i\right)p^*\left(w_k^i\right)dw_k^i =$$

$$= \frac{2(1-\varepsilon)\sqrt{2\pi}}{\sqrt{2\pi}} \int_0^{k_\varepsilon^i} e - \frac{1}{2}\left(\frac{w_k^i}{\sigma_{1i}}\right)^2 d\left(\frac{w_k^i}{\sigma_{1i}}\right) = 2(1-\varepsilon)\phi_L^i\left(k_\varepsilon^i\right) \quad (63)$$

The function $\phi_L^i(\cdot)$ is a Laplace function for which exists table of values. Matrix $\boldsymbol{M}$ has a form

$$\boldsymbol{M} = \begin{bmatrix} m_1 & & 0 \\ & \ddots & \\ 0 & & m_p \end{bmatrix} \quad (64)$$

and second derivative of functional, based on (59)

$$\nabla_{\theta}^{2} J^{a}(\boldsymbol{\theta}) = trE\left\{\left(\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^T\right)\boldsymbol{M}\right\} = E\left\{\boldsymbol{\varphi}_k^T \boldsymbol{M}\boldsymbol{\varphi}_k\right\} \quad (65)$$

where is $J^{ph}(\boldsymbol{\theta}) \cong J^{a}(\boldsymbol{\theta})$

Empirical functional of $\nabla_{\theta}^{2} J^{a}(\boldsymbol{\theta})$ is

$$k\nabla_{\theta}^{2} J^{a}(\boldsymbol{\theta}_{k-1}) = \sum_{i=1}^{k} \boldsymbol{\varphi}_k^T \boldsymbol{M}\boldsymbol{\varphi}_k \quad (66)$$

As for algorithm (51) and (57) is obtained

$$\boldsymbol{P}_k^a = \boldsymbol{P}_{k-1}^a - \boldsymbol{P}_{k-1}^a\boldsymbol{\varphi}_k^T\left[\boldsymbol{\varphi}_k\boldsymbol{P}_{k-1}\boldsymbol{\varphi}_k^T + \boldsymbol{M}^{-1}\right]^{-1}\varphi_k\boldsymbol{P}_k^a \quad (67)$$

where

$$\boldsymbol{P}_k^a = \left[k\nabla_{\theta}^{2} J_k^{a}(\boldsymbol{\theta}_{k-1})\right]^{-1} \quad (68)$$

Similarly as for relation (51) for parameter estimate we have next relation

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{P}_k^a \boldsymbol{\varphi}_k^T \boldsymbol{\psi}(\boldsymbol{e}_k) \qquad (69)$$

Relation (67) and (69) represent second form of recursive algorithm. Such algorithm is considered in [21]. The explicit form of the algorithm considered in this paper is

Robust algorithm

- Tukey class of distributions

$$P = (1-\varepsilon)N(0,\boldsymbol{\Sigma}_1) + \varepsilon N(0-\boldsymbol{\Sigma}_2) \quad , \ \varepsilon \in [0,1)$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_{11}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{1p}^2 \end{bmatrix} \ , \ \boldsymbol{\Sigma}_2 = \begin{bmatrix} \sigma_{21}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{2p}^2 \end{bmatrix}$$

$$\sigma_{1i}^2 \ll \sigma_{2i}^2 \quad , \quad (i = 1,2,...p)$$

- Prediction error

$$\boldsymbol{e}_k = \boldsymbol{y}_k - \boldsymbol{\varphi}_k \hat{\boldsymbol{\theta}}_{k-1}$$

- Matrix gain

$$\boldsymbol{P}_k = \boldsymbol{P}_{k-1} - \boldsymbol{P}_{k-1}\boldsymbol{\varphi}_k^T \left[\boldsymbol{\varphi}_k \boldsymbol{P}_{k-1}\boldsymbol{\varphi}_k^T + \boldsymbol{N}^{-1}(\boldsymbol{e}_k)\right]^{-1} \boldsymbol{\varphi}_k \boldsymbol{P}_k \ , \ \boldsymbol{P}_0 = g\boldsymbol{I} \ , \ g \gg 1$$

$$\boldsymbol{N}_a(\boldsymbol{e}_k) = \begin{bmatrix} \psi_{ph}^1 (e_k^1)' & & 0 \\ & \ddots & \\ 0 & & \psi_{ph}^p (e_k^p)' \end{bmatrix}$$

$$\left(\psi_{ph}^i (e_k^i)\right)' = \frac{(k_\varepsilon^i)^3}{\left(\sqrt{(k_\varepsilon^i)^2 + (e_k^i)^2}\right)^{3/2}}$$

$$i = 1,2,...,p$$

- Parameter estimate

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{P}_k \boldsymbol{\varphi}_k^T \boldsymbol{\psi}(\boldsymbol{e}_k) \ , \ \boldsymbol{\theta}_0 = \boldsymbol{0}$$

$$\boldsymbol{\psi}(\boldsymbol{e}_k) = \begin{bmatrix} \psi^1(e_k^1) \\ \psi^2(e_k^2) \\ \vdots \\ \psi^p(e_k^p) \end{bmatrix}$$

$$\psi^i(e_k^i) = \begin{cases} e_k^i & , \ |e_k^i| \le k_\varepsilon^i \\ k_\varepsilon^i \mathrm{sign} e_k^i & , \ |e_k^i| > k_\varepsilon^i \end{cases}$$

$$i = 1,2,...,p$$

V. CONVERGENCEANALYSIS

The convergence analysis of the algorithm (50) and (57) is based on martingale theory [27]. We assume that $\{\boldsymbol{w}_k\}$ is a martingale- difference sequence with respect to an increasing sequence of σ - fields $\{F_k : k \in Z_+\}$ defined on the underlying probability space (Ω, F, P). We shell require the following conditions to hold.

A) Hypotheses for stochasic disturbance

(A1) $\{\boldsymbol{w}_k\}$ is a sequence of martingale difference with symmetric probability distributionfunction

(A2) $\sup_k E\left\{\|\boldsymbol{w}_k\|^2 \middle| F_{k-1}\right\} = \sigma^2 < \infty \ w.p.1$

(A3)The zeros of $\det \boldsymbol{C}\left(q^{-1}\right)$ lie outside the closed unit disc

B) Hypotheses for nonlinearfunction $\boldsymbol{\psi}(\cdot)$

(B1) The function $\boldsymbol{\psi}(\cdot)$ is odd and continuous everywhere

(B2) The function $\boldsymbol{\psi}(\cdot)$ is uniformly bounded

C) Hypotheses for pseudo - Huberfunction

(C1) The function $\left(\psi^i_{ph}(\cdot)\right)'$, $i = 1,2,...,p$ are even bounded functions

(C2) Elements of matrix $\boldsymbol{N}\left(\boldsymbol{e}_k\right)$ is uniformly bounded

$$\boldsymbol{N}\left(\boldsymbol{e}_k\right) = \begin{bmatrix} \psi^1_{ph}\left(e^1_k\right)' & & 0 \\ & \ddots & \\ 0 & & \psi^p_{ph}\left(e^p_k\right)' \end{bmatrix}$$

satisfay the condition $\left(\psi^i_{ph}\left(e^i_k\right)\right)' \in [0, \infty)$, $i = 1,2,...,p$

(C3) $E\left\{\boldsymbol{N}\left(\boldsymbol{e}_k\right) \middle| F_{k-1}\right\} = \boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)$, $\tilde{\boldsymbol{\theta}}_k = \tilde{\boldsymbol{\theta}}_{k-\theta}$

$$\boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right) = \begin{bmatrix} \eta^1\left(\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)^1\right) & 0 \\ 0 & \eta^p\left(\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)^p\right) \end{bmatrix}$$

where $\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)^i$, $i = 1,2,...,p$ is the i-th row of the matrix $\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}$ and

$$\eta^i\left(\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)^i\right) = E\left\{\psi^i_{ph}\left(e^i_k\right) \middle| F_{k-1}\right\}, \quad i = 1,2,...,p$$

$$\eta^i(\cdot) \in (0, \infty)$$

D) Hypotheses for conditional mathematical exception for trace of matrix gain P_k

$$r_k^a = E\left\{tr\boldsymbol{P}_k^{-1}\middle|F_{k-1}\right\}$$

(D1) $$r_k^a = r_{k-1}^a + tr\left\{\boldsymbol{\varphi}_k^T\boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)\boldsymbol{\varphi}_k\right\}$$

$$\liminf_{k\to\infty} r_k^a = \infty \quad , \quad w.p.1$$

E) Hypotheses for generalized strictly positive real conditions

(E1) There exists the strictly passive operator **H** such that

$$\boldsymbol{H}\boldsymbol{z}_k = \boldsymbol{\Phi}_1\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{z}_k\right) - \frac{1}{2}\boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{z}_k\right)\boldsymbol{z}_k$$

where $\boldsymbol{\Phi}_1\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{z}_k\right) = E\left\{\boldsymbol{\psi}\left(-\boldsymbol{e}_k\right)\middle|F_{k-1}\right\}$, $\boldsymbol{z}_k = \boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}$

F) Hypotheses about the persistent excitation condition (F1). There exist constans $c > 1$, $k_{\psi p} > 1$ such that

$$\lim_{k\to\infty}\frac{\log^c\left(k_{\psi p} r_k^a\right)}{\lambda_{\min}\left\{\boldsymbol{P}_k^{-1}\right\}} = 0$$

Now we will prove following lemma.

Lemma: Let us consider the model (18) – (21) and algorithm (50) and (57) subject to the assumption (A3), (C1) – (C3) and (D1). Then

$$\sum_{k=1}^{\infty}\frac{tr\left(\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\right)}{\log^c\left(k_{\psi p} r_k^a\right)} < \infty \quad , \quad c > 1 \quad , \quad k_{\psi p} > 1 \quad , \quad w.p.1$$

Proof: From relation (53) we have

$$\boldsymbol{P}_{k-1}^{-1} = \boldsymbol{P}_k^{-1} - \boldsymbol{\varphi}_k^T\boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k \qquad (70)$$

Let $\boldsymbol{\varphi}_{ki}$ is the i-th row of the matrix $\boldsymbol{\varphi}_k$. Using relation (70) it follows that

$$\boldsymbol{P}_{k-1}^{-1} \le \boldsymbol{P}_k^{-1} - \left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{\varphi}_{ki}^T\boldsymbol{\varphi}_{ki} = \boldsymbol{P}_k^{-1}\left(I - \left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{P}_k\boldsymbol{\varphi}_{ki}^T\boldsymbol{\varphi}_{ki}\right) \qquad (71)$$

From relation (71) one can get

$$\det\boldsymbol{P}_{k-1}^{-1} \le \det\boldsymbol{P}_k^{-1}\det\left(\boldsymbol{I} - \left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{P}_k\boldsymbol{\varphi}_{ki}^T\boldsymbol{\varphi}_{ki}\right) \qquad (72)$$

The eigenvalues of matrix $\left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{P}_k\boldsymbol{\varphi}_{ki}^T\boldsymbol{\varphi}_{ki}$ are solution of equation

$$\det\left(\left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{P}_k\boldsymbol{\varphi}_{ki}^T\boldsymbol{\varphi}_{ki} - \lambda\boldsymbol{I}\right) = 0 \qquad (73)$$

The solutions of equations are

$$\lambda_1 = \left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{\varphi}_{ki}^T\boldsymbol{P}_k\boldsymbol{\varphi}_{ki} \ , \ \lambda_2 = \lambda_3 = \cdot\cdot = \lambda_p = 0 \qquad (74)$$

Owing the relation (74) we have

$$\det\left(I - \left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{P}_k\boldsymbol{\varphi}_{ki}^T\boldsymbol{\varphi}_{ki}\right) = 1 - \left(\psi_{ph}^{i}\left(e_k^i\right)\right)'\boldsymbol{\varphi}_{ki}^T\boldsymbol{P}_k\boldsymbol{\varphi}_{ki} \qquad (75)$$

By using relations (72) and (75) it follows that

$$\left(\psi_{ph}^{i}\left(e_k^i\right)\right)' \boldsymbol{\varphi}_{ki}^T \boldsymbol{P}_k \boldsymbol{\varphi}_{ki} \le \frac{\det \boldsymbol{P}_k^{-1} - \det \boldsymbol{P}_{k-1}^{-1}}{\det \boldsymbol{P}_k^{-1}} \tag{76}$$

Let us consider

$$r_k = tr\left\{\boldsymbol{P}_k^{-1}\right\} = tr\left\{\sum_{i=1}^{k} \boldsymbol{\varphi}_k^T \boldsymbol{N}\left(\boldsymbol{e}_i\right)\boldsymbol{\varphi}_i\right\} \tag{77}$$

Prediction error has a form

$$\boldsymbol{e}_k = \boldsymbol{y}_k - \boldsymbol{\varphi}_k \tilde{\boldsymbol{\theta}}_{k-1} = -\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_k \tilde{\boldsymbol{\theta}}_{k-1} + \boldsymbol{w}_k \tag{78}$$

From relation (77) and conditions of lemma (C3) and (D1) one can get

$$E\left\{r_k \left| F_{k-1}\right.\right\} = tr\left\{\sum_{i=1}^{k} \boldsymbol{\varphi}_i^T E\left\{\boldsymbol{N}\left(\boldsymbol{e}_i\right)\right\}\boldsymbol{\varphi}_i\right\} = tr\left\{\sum_{i=1}^{k} \boldsymbol{\varphi}_k^T \boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_i \tilde{\boldsymbol{\theta}}_{i-1}\right)\boldsymbol{\varphi}_i\right\} = r_k^a \tag{79}$$

Using assumption (D1) it follows that

$$\liminf_{n\to\infty} \ \mathrm{E}\left\{r_k \left| F_{k-1}\right.\right\} = \infty \tag{80}$$

and consequently

$$\liminf_{k\to\infty} r_k = \infty \ \ , \ \ w.p.1 \tag{81}$$

Based on condition (C1) – (C3) it follows that it can be found $\exists k_{\psi p} > 1$ for which

$$k_{\psi p}\eta^i\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)^{-1}\boldsymbol{\varphi}_k \tilde{\boldsymbol{\theta}}_{k-1}\right)^i \ge \left(\psi_{ph}\left(e_k^i\right)\right)' \ , \quad i = 1,2,...,p \tag{82}$$

From last relation it follows that

$$k_{\psi p}\boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_i \tilde{\boldsymbol{\theta}}_{i-1}\right) \ge \boldsymbol{N}\left(\boldsymbol{e}_k\right) \tag{83}$$

Using relations (79) and (83) we have

$$r_k \le k_{\psi p} tr\left\{\sum_{i=1}^{k} \boldsymbol{\varphi}_k^T \boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}\left(q^{-1}\right)\boldsymbol{\varphi}_i \tilde{\boldsymbol{\theta}}_{i-1}\right)\boldsymbol{\varphi}_i\right\} = k_{\psi p} r_k^a \tag{84}$$

Let matrix P_k has d x d dimensions. Let us notice that

$$\det \boldsymbol{P}_k^{-1} = \prod_{k=1}^{d} \lambda_i\left\{\boldsymbol{P}_k^{-1}\right\} \le \lambda_{\max}^d\left\{\boldsymbol{P}_k^{-1}\right\} \tag{85}$$

Further we have

$$r_k = \sum_{k=1}^{d} \lambda_k\left\{\boldsymbol{P}_k^{-1}\right\} \ge \lambda_{\max}\left\{\boldsymbol{P}_k^{-1}\right\} \tag{86}$$

where $\lambda_k\{\cdot\}$ is eigenvalue, with $\lambda_{\max}\{\cdot\}$ being the maximal eigenvalue
From last two relations it follows that

$$r_k \ge \left(\det \boldsymbol{P}_k^{-1}\right)^{1/d} \tag{87}$$

Using relation (84) it follows that

$$\sum_{k=k_0}^{\infty} \frac{tr\left\{\boldsymbol{\varphi}_k \boldsymbol{P}_k \boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k^T\right\}}{\log^c\left(k_{\psi p} r_k^a\right)} \le \sum_{k=k_0}^{\infty} \frac{tr\left\{\boldsymbol{\varphi}_k \boldsymbol{P}_k \boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k^T\right\}}{\log^c r_k}$$

From relations (76) and (87) we have

$$\sum_{k=k_0}^{\infty}\frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k^T\right\}}{\log^c r_k}=\sum_{k=k_0}^{\infty}\sum_{i=1}^{d}\frac{\left(\psi_{ph}^i\left(e_k^i\right)\right)'\left\{\boldsymbol{\varphi}_{ki}\boldsymbol{P}_k\boldsymbol{\varphi}_{ki}^T\right\}}{\log^c r_k}\leq$$

$$d^c\sum_{i=1}^{p}\sum_{k=k_0}^{\infty}\frac{\det\boldsymbol{P}_k^{-1}-\det\boldsymbol{P}_{k-1}^{-1}}{\det\boldsymbol{P}_k^{-1}\log^c\left(\det\boldsymbol{P}_k^{-1}\right)}=d^c\sum_{i=1}^{p}\sum_{k=k_0}^{\infty}\int_{\det P_{k_0}^{-1}}^{\det P_k^{-1}}\frac{dx}{\log^c x}$$

$$=\frac{pd^c}{\left(c-1\right)\log^c\left(\det\boldsymbol{P}_{k_0}^{-1}\right)}<\infty\ \ ,\ w.p.1 \tag{88}$$

In relation (88) two facts are used

$$\det\boldsymbol{P}_k^{-1}>1\ \ ,\ \text{for } \forall k\geq k_0 \tag{89}$$

$$\det\boldsymbol{P}_\infty^{-1}=\infty\ \ ,\ w.p.1 \tag{90}$$

From relation (47) one can get

$$\sup\left(\psi_{ph}^i\left(e_k^i\right)\right)'=1\ \ ,\quad i=1,2,...,p \tag{91}$$

For that case

$$\boldsymbol{N}\left(\boldsymbol{e}_k\right)=\begin{bmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{bmatrix} \tag{92}$$

By using condition (C2) and relation (92) it is possible to conclude that

$$\boldsymbol{0}\leq\boldsymbol{N}\left(\boldsymbol{e}_k\right)\leq\boldsymbol{I}$$

Let us define

$$S=\left\{\boldsymbol{0}\leq\boldsymbol{N}\left(\boldsymbol{e}_k\right)\leq\boldsymbol{I}\right\}\ ,\ S_1=\left\{\boldsymbol{0}\leq\boldsymbol{N}\left(\boldsymbol{e}_k\right)<\boldsymbol{I}\right\}\ \ ,\ S_2=\left\{\boldsymbol{N}\left(\boldsymbol{e}_k\right)=\boldsymbol{I}\ \right\}$$

$$S=S_1\cup S_2$$

Now from (88) it follows that

$$\sum_{k=k_0,S}^{\infty}\frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k^T\right\}}{\log^c r_k}=\sum_{k=k_0,S_1}^{\infty}\frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k^T\right\}}{\log^c r_k}+$$

$$+\sum_{k=k_0,S_2}^{\infty}\frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k^T\right\}}{\log^c r_k}<\infty$$

From last relation we have

$$\sum_{k=k_0,S_2}^{\infty}\frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{N}\left(\boldsymbol{e}_k\right)\boldsymbol{\varphi}_k^T\right\}}{\log^c r_k}=\sum_{k=k_0}^{\infty}\frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\right\}}{\log^c r_k}<\infty \tag{93}$$

From relations (84) and (93) follows proof of lemma ▪

Now we will formulate main result of the paper.

Theorem: Consider the model (18) – (21) and the algorithms (50) and (57) subject to the assumptions of the lemma and assume further that the following hypotheses are satisfied : (A1) – (A2), (B1) – (B2), (E1) and (F1) Then

$$P\left\{\lim_{k\to\infty}\hat{\boldsymbol{\theta}}_k=\boldsymbol{\theta}\right\}=1$$

Proof: Introducing the Lyapunov stochastic function $V_k=\tilde{\boldsymbol{\theta}}_k^T\boldsymbol{P}_k\tilde{\boldsymbol{\theta}}_k$ we obtain using (51)

$$V_k = \left[\tilde{\boldsymbol{\theta}}_{k-1} + \boldsymbol{P}_k\boldsymbol{\varphi}_k^T\boldsymbol{\psi}(\boldsymbol{e}_k)\right]^T \boldsymbol{P}_k^{-1}\left[\tilde{\boldsymbol{\theta}}_{k-1} + \boldsymbol{P}_k\boldsymbol{\varphi}_k^T\boldsymbol{\psi}(\boldsymbol{e}_k)\right] =$$
$$\tilde{\boldsymbol{\theta}}_{k-1}^T\boldsymbol{P}_k^{-1}\tilde{\boldsymbol{\theta}}_{k-1} + 2\tilde{\boldsymbol{\theta}}_{k-1}^T\boldsymbol{\varphi}_k^T\boldsymbol{\psi}(\boldsymbol{e}_k) + \boldsymbol{\psi}^T(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\boldsymbol{\psi}(\boldsymbol{e}_k) \qquad (94)$$

From relation (53) we have

$$\boldsymbol{P}_k^{-1} = \boldsymbol{P}_{k-1}^{-1} + \boldsymbol{\varphi}_k^T\boldsymbol{N}(\boldsymbol{e}_k)\boldsymbol{\varphi}_k \qquad (95)$$

From last relation it follows that

$$\tilde{\boldsymbol{\theta}}_{k-1}^T P_k^{-1}\tilde{\theta}_{k-1} = \tilde{\boldsymbol{\theta}}_{k-1}^T\left[\boldsymbol{P}_{k-1}^{-1} + \boldsymbol{\varphi}_k^T\boldsymbol{N}(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\right]\tilde{\boldsymbol{\theta}}_{k-1} = V_{k-1} + \left(\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}^T\right)^T\boldsymbol{N}(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1} \qquad (96)$$

Using relations (94) and (96) one concludes that

$$V_k = V_{k-1} + \left(\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}^T\right)^T\boldsymbol{N}(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1} + 2\left(\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}^T\right)^T\boldsymbol{\psi}(\boldsymbol{e}_k) + \boldsymbol{\psi}^T(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\boldsymbol{\psi}(\boldsymbol{e}_k) \qquad (97)$$

Using conditions (A1) – (A2) , (B1) and (D1) of theorem it follows that

$$E\left\{\frac{V_k}{\log^c\left(k_{\psi p}r_k^a\right)}\middle|F_{k-1}\right\} = \frac{V_k}{\log^c\left(k_{\psi p}r_k^a\right)} + \frac{\left(\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)^T\boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}(q^{-1})\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}}{\log^c\left(k_{\psi p}r_k^a\right)} -$$
$$-\frac{2\left(\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)^T\boldsymbol{\Phi}_1\left(\boldsymbol{C}^{-1}(q^{-1})\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)}{\log^c\left(k_{\psi p}r_k^a\right)} + E\left\{\frac{\boldsymbol{\psi}^T(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\boldsymbol{\psi}(\boldsymbol{e}_k)}{\log^c\left(k_{\psi p}r_k^a\right)}\middle|F_{k-1}\right\} \qquad (98)$$

From condition (B2) of theorem we have

$$\boldsymbol{\psi}^T(\boldsymbol{e}_k)\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\boldsymbol{\psi}(\boldsymbol{e}_k) \le \left\|\boldsymbol{\psi}(\boldsymbol{e}_k)\right\|^2\lambda_{\max}\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\right\} \le k_\psi tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\right\} \qquad (99)$$

Where

$$k_\psi \in \left(\sup_k\left\|\boldsymbol{\psi}(\boldsymbol{e}_k)\right\|^2, \infty\right) \qquad (100)$$

Now from relations (98) and (99) it follows that

$$E\left\{\frac{V_k}{\log^c\left(k_{\psi p}r_k^a\right)}\middle|F_{k-1}\right\} = \frac{V_k}{\log^c\left(k_{\psi p}r_k^a\right)} - \frac{2\left(\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)^T}{\log^c\left(k_{\psi p}r_k^a\right)}\cdot$$
$$\left[\boldsymbol{\Phi}_1\left(\boldsymbol{C}^{-1}(q^{-1})\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right) - \frac{1}{2}\boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}(q^{-1})\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right] + k_\psi\frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\right\}}{\log^c\left(k_{\psi p}r_k^a\right)} \qquad (101)$$

By using assumption (E1) of theorem and property of passive operators we have

$$S_k = 2\sum_{i=1}^{k}\left(\boldsymbol{\varphi}_i\tilde{\boldsymbol{\theta}}_{i-1}\right)^T\left[\boldsymbol{\Phi}_1\left(\boldsymbol{C}^{-1}(q^{-1})\boldsymbol{\varphi}_i\tilde{\boldsymbol{\theta}}_{i-1}\right) - \frac{1}{2}\boldsymbol{\Phi}_{\psi p}\left(\boldsymbol{C}^{-1}(q^{-1})\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right)\boldsymbol{\varphi}_k\tilde{\boldsymbol{\theta}}_{k-1}\right] + k_n \ge 0 ,$$

$$k_n \in [0, \infty) \qquad (102)$$

Let us define the quantity

$$T_k = R_k + \frac{S_k}{\log^c\left(k_{\psi p}r_k^a\right)} \ , \ R_k = \frac{V_k}{\log^c\left(k_{\psi p}r_k^a\right)} \qquad (103)$$

Using (101) – (103) one concludes

$$E\left\{T_k\middle|F_{k-1}\right\} = T_{k-1} + \frac{tr\left\{\boldsymbol{\varphi}_k\boldsymbol{P}_k\boldsymbol{\varphi}_k^T\right\}}{\log^c\left(k_{\psi p}r_k^a\right)} \qquad (104)$$

From result of Lemma and relation (104) it follows that

$$\lim_{n\to\infty} T_k = T^* \ , \ w.p.1 \qquad (105)$$

By using relation (105) we have $\lim_{n\to\infty} R_k = R^*$, $w.p.1$ (T^* and R^* are finite constants) and one can get

$$R_k = \frac{tr\left\{\boldsymbol{P}_k^{-1}\tilde{\boldsymbol{\theta}}_k\tilde{\boldsymbol{\theta}}_k^T\right\}}{\log^c\left(k_{\psi p}r_k^a\right)} \le \frac{\lambda_{\min}\left\{\boldsymbol{P}_k^{-1}\right\}\left\|\tilde{\boldsymbol{\theta}}_k\right\|^2}{\log^c\left(k_{\psi p}r_k^a\right)} = \frac{\left\|\tilde{\boldsymbol{\theta}}_k\right\|^2}{\dfrac{\log^c\left(k_{\psi p}r_k^a\right)}{\lambda_{\min}\left\{\boldsymbol{P}_k^{-1}\right\}}} \quad (106)$$

From assumption (E1), relations (103) and (105) and last relation follows the proof of theorem ▪

VI. SIMULATION STUDY

In this paper we consider properties of robust recursive algorithm, for estimation of MIMO NARMAX model, on the simulation level. In what follows we will consider two inputs and two outputs model

$$\boldsymbol{A}\left(q^{-1}\right)\boldsymbol{y}_k = \boldsymbol{B}\left(q^{-1}\right)\boldsymbol{f}\left(\boldsymbol{u}_k\right) + \boldsymbol{C}\left(q^{-1}\right)\boldsymbol{w}_k$$

where

$$\boldsymbol{A}\left(q^{-1}\right) = \boldsymbol{I} + \boldsymbol{A}_1 q^{-1} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + \begin{bmatrix}0.3 & 0.4\\ -0.65 & 0.7\end{bmatrix}q^{-1}$$

$$\boldsymbol{B}\left(q^{-1}\right) = \boldsymbol{B}_1\left(q^{-1}\right) = \begin{bmatrix}0.8 & -0.5\\ -0.35 & 0.95\end{bmatrix}q^{-1}$$

$$\boldsymbol{C}\left(q^{-1}\right) = \boldsymbol{I} + \boldsymbol{C}_1\left(q^{-1}\right) = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + \begin{bmatrix}0.58 & 0.41\\ -0.75 & 0.85\end{bmatrix}q^{-1}$$

The $\boldsymbol{D}_i$, $i = 1,2,3$ matrices have a form

$$\boldsymbol{D}_1 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} ,\boldsymbol{D}_2 = \begin{bmatrix}0 & 0\\ 0 & 0.5\end{bmatrix} ,\boldsymbol{D}_3 = \begin{bmatrix}0.35 & 0\\ 0 & 0.5\end{bmatrix}$$

The vectors $\boldsymbol{\Gamma}_i\left(\boldsymbol{u}_k\right)$, $i = 1,2,3$ are

$$\boldsymbol{\Gamma}_1\left(\boldsymbol{u}_k\right) = \begin{bmatrix}u_k^1\\ u_k^2\end{bmatrix} , \boldsymbol{\Gamma}_2\left(\boldsymbol{u}_k\right) = \begin{bmatrix}\left(u_k^1\right)^2\\ \left(u_k^2\right)^2\end{bmatrix} , \boldsymbol{\Gamma}_3\left(\boldsymbol{u}_k\right) = \begin{bmatrix}\left(u_k^1\right)^3\\ \left(u_k^2\right)^3\end{bmatrix}$$

The nonlinear black is

$$\boldsymbol{f}\left(\boldsymbol{u}_k\right) = \sum_{i=1}^{3}\boldsymbol{D}_3\boldsymbol{\Gamma}_3\left(\boldsymbol{u}_k\right) = \begin{bmatrix}u_k^1 & +0.35\left(u_k^1\right)^3\\ u_k^2 & +0.5\left(u_k^2\right)^2 + 0.5\left(u_k^2\right)^3\end{bmatrix}$$

and unknown parameter

$$\left(\boldsymbol{\theta}^M\right)^T = \left[\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{B}_1\boldsymbol{D}_2, \boldsymbol{B}_1\boldsymbol{D}_3, \boldsymbol{C}_1\right]$$

By using above matrices one can get that parameter

$$\left(\boldsymbol{\theta}^M\right)^T = \begin{bmatrix}0.3 & 0.4 & 0.8 & -0.5 & 0 & -0.25 & 0.28 & -0.25 & 0.58 & 0.41\\ -0.65 & 0.7 & -0.35 & 0.95 & 0 & 0.475 & -0.123 & 0.475 & -0.75 & 0.85\end{bmatrix}$$

For our example

$$\boldsymbol{\varphi}_k = \begin{bmatrix}\boldsymbol{x}_k^T & 0\\ 0 & \boldsymbol{x}_k^T\end{bmatrix}$$

where

$$\boldsymbol{x}_k^T = \left[\boldsymbol{y}_{k-1}^T, \boldsymbol{\Gamma}_1^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\Gamma}_2^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{\Gamma}_3^T\left(\boldsymbol{u}_{k-1}\right), \boldsymbol{e}_{k-1}^T\right]$$

$$\boldsymbol{e}_k = \boldsymbol{y}_k - \boldsymbol{\varphi}_k \boldsymbol{\theta}_{k-1}$$

The components of input signal $\boldsymbol{u}_k = \left[u_k^1, u_k^2\right]$ are random variables which are uniformly distributed in an interval (-2 , 2).
The stochastic disturbance has non – Gaussian distribution

$$w_k^i = \left(1-\varepsilon_i\right) N\left(0,\sigma_{1i}^2\right) + \varepsilon_i N\left(0,\sigma_{2i}^2\right) \quad , \quad i = 1,2$$

where $N\left(m,\sigma^2\right)$ is Gaussian distribution with mean *m* and variance σ^2. In the simulations is used

$$\sigma_{1i}^2 = 1 \quad , \quad \sigma_{2i}^2 = 100 \quad , \quad i = 1,2$$

$$\varepsilon_i = \varepsilon \quad , \quad i = 1,2$$

In all cases it is assumed that the parameter of Huber function

$$k_\varepsilon^i = 3 \quad , \quad i = 1,2$$

Parameters $\left(\boldsymbol{D}_i\right)_k$, $i = 1,2,\ldots s$, for known matrix $\boldsymbol{D}_1$ as supposed in the paper, can be determined in the following way

$$\left(\boldsymbol{D}_2\right)_k = \left(\boldsymbol{B}_1^{-1}\right)_k \left(\boldsymbol{B}_1^{-1}\boldsymbol{D}_2\right)_k$$

$$\left(\boldsymbol{D}_s\right)_k = \left(\boldsymbol{B}_1^{-1}\right)_k \left(\boldsymbol{B}_1\boldsymbol{D}_s\right)_k$$

In the our example it is necessary to determine, according with form $\left(\boldsymbol{\theta}^M\right)^T$, matrices $\boldsymbol{D}_2$ and $\boldsymbol{D}_3$. When $r \neq p$ (model with *r* – inputs and *p* – outputs) the matrix $\boldsymbol{B}_1$ is rectangular and it is necessary to use a pseudo inverse matrix $\boldsymbol{B}_1^+$. For rectangular matrix $\boldsymbol{B}_1$ with *p* x *r* dimensions $p \neq r$ the pseudo inverse matrix is

$$\boldsymbol{B}_1^+ = \left(\boldsymbol{B}_1^T \boldsymbol{B}_1\right)^{-1} \boldsymbol{B}_1^T$$

Now matrices can be determined as

$$\left(\boldsymbol{D}_2\right)_k = \left(\boldsymbol{B}_1\right)_k^+ \left(\boldsymbol{B}_1\boldsymbol{D}_2\right)_k$$

$$\left(\boldsymbol{D}_s\right)_k = \left(\boldsymbol{B}_1\right)_k^+ \left(\boldsymbol{B}_1\boldsymbol{D}_s\right)_k$$

In this paper we compare two algorithms. The first one is algorithm designed for the case when disturbance has Gaussian distribution (linear algorithm). The second algorithm is designed for case of non – Gaussian distribution of disturbance (that is algorithmproposed in this paper.
The linear extended least squares algorithm (ELS) is

$$\boldsymbol{\theta}_k^G = \boldsymbol{\theta}_{k-1}^G + \boldsymbol{P}_k^G \left(\boldsymbol{\varphi}_k^G\right)^T \boldsymbol{e}_k^G \quad , \quad \boldsymbol{e}_k^G = \boldsymbol{y}_k - \boldsymbol{\varphi}_k^G \boldsymbol{\theta}_{k-1}^G$$

$$\boldsymbol{P}_k^G = \boldsymbol{P}_{k-1}^G - \boldsymbol{P}_{k-1}^G \left(\boldsymbol{\varphi}_k^G\right)^T \left[\boldsymbol{\varphi}_k^G \boldsymbol{P}_{k-1}^G \left(\boldsymbol{e}_k^G\right)^T + \boldsymbol{I}\right] \boldsymbol{\varphi}_k^G \boldsymbol{P}_k^G$$

The algorithm proposed in this paper (robust extended least square- RELS) is

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{P}_k \boldsymbol{\varphi}_k^T \boldsymbol{\psi}\left(\boldsymbol{e}_k\right)$$

$$\boldsymbol{P}_k = \boldsymbol{P}_{k-1} - \boldsymbol{P}_{k-1}\boldsymbol{\varphi}_k^T \left[\boldsymbol{\varphi}_k \boldsymbol{P}_{k-1} \boldsymbol{\varphi}_k^T + \boldsymbol{N}^{-1}\left(\boldsymbol{e}_k\right)\right]^{-1} \boldsymbol{\varphi}_k \boldsymbol{P}_{k-1}$$

The estimation error E_k^G for linear algorithm (extended least squares) is

$$\mathrm{E}_k^G = \log\left\|\boldsymbol{\theta}_k^G - \boldsymbol{\theta}\right\|^2$$

And for algorithm in this paper (robust extended least squares)

$$\mathrm{E}_k = \log\left\|\boldsymbol{\theta}_k - \boldsymbol{\theta}\right\|^2$$

In the next figures (Fig.1, Fig.2) the comparison between linear and robust algorithm is presented. From figures it is possible to see that behaviour of robust recursivealgorithm is superior in comparison to linear algorithm. The superiority increases with increase in contamination degree

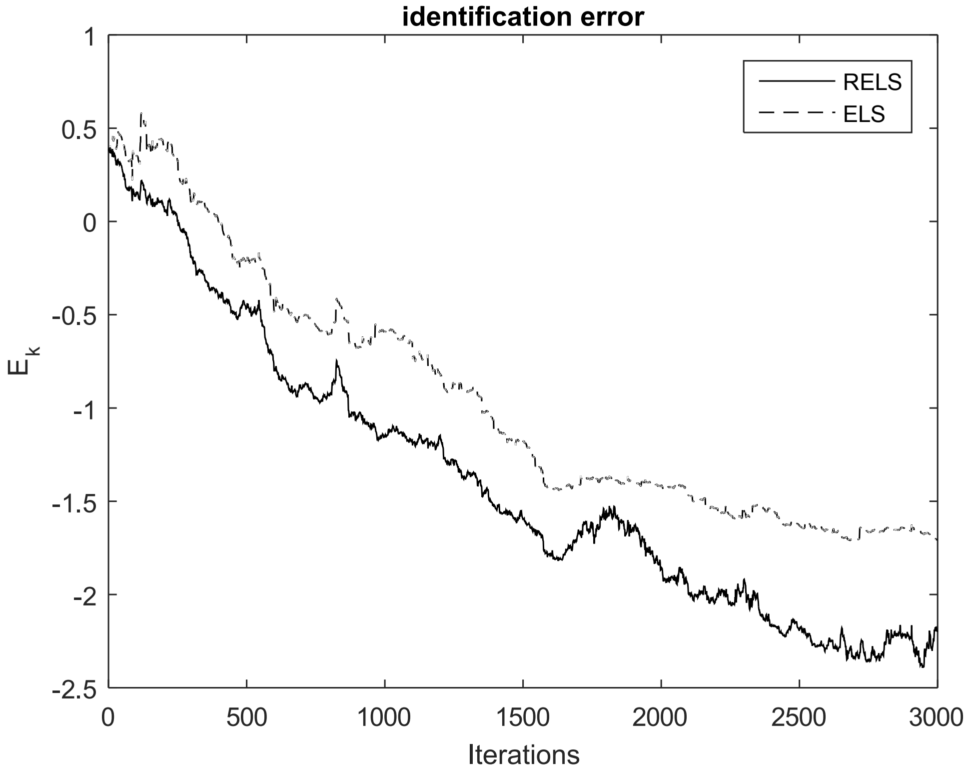

Fig.1 Comparison of robust and linear algorithms for $\varepsilon = 0.05$

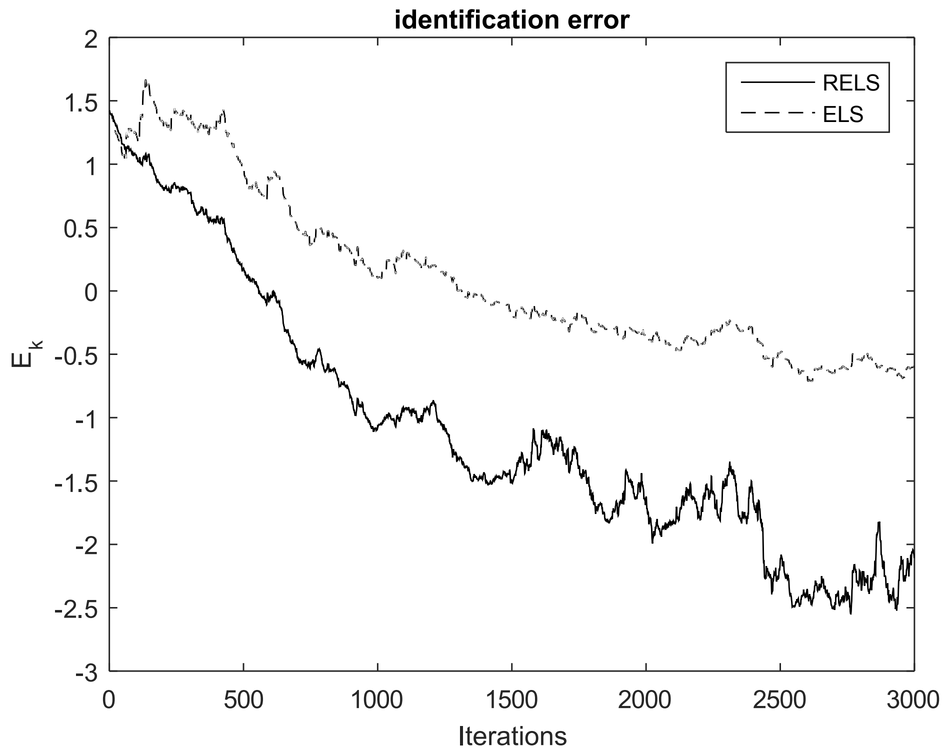

Fig.2 Comparison of robust and linear algorithms for $\varepsilon = 0.2$

VII.

VIII. CONCLUSION

The paper considers design of recursive algorithm by using Newton – Raphson method. It is, also, assumed that observations have outliers. The key ingredient in systems identification is identification criterion (loss function) and for Newton – Raphson method loss function must be second order differentiable. Huber loss function has only first derivative. Owing that fact it is introduced pseudo – Huber loss function which has derivatives of all degrees and which behaves similarly as Huber loss function. The recursive algorithm is based on both functions. The first derivative in method is based on Huber loss function while second derivative is based on pseudo – Huber loss function. The convergence analysis is performed for given robust recursive algorithm. Further investigation will be related to identification of time varying nonlinear systems.

REFERENCES

[1]. S. Skogestad, I. Postelethwaite,*Multivariable Feedback Control. Analysis and Design, second ed.*, (Wiley, New York, 2005)

[2]. S. Fan, F. Ding, T. Hayat, Recursive identification of errors-in-variables systems based on the correlation analysis. Circuits, Systems, and Signal Processing **39** (12), 5951-5981 (2020). https://doi.org/10.1007/s00034-020-01441-7.

[3]. Y. Pu, Y. Yang, J. Chen, Some stochastic gradient algorithms for Hammerstein systems with piecewise linearity, Circuits, Systems, and Signal Processing **40** (4),1635-1651 (2021).https://doi.org/10.1007/s00034-020-01554-z

[4]. A. Pouradabi, A. Rastegarnia, S. Zandi, W.M. Bazzi, S. Sanel, A class of diffusion proportionate subband adaptive filters for sparse system identification over distributed networks. Circuits, Systems, and Signal Processing(2021). https:/doi.org./10.1007/s00034-021-01766-x.

[5]. P. Caines,*Stochastic Systems.*(Wiley, New York, 1988)

[6]. Y. Zhu,*Multivariable System Identification for Process Control.*(Elsevier Science, London, 2000)

[7]. R. Pearson,*Exploring Data in Engineering, the Science and Medicine.*(Oxford University Press, Oxford,2011)

[8]. P. Huber, Robust estimation of a location parameter. Annals of Mathematical Statistics.**35** (1), 73-101 (1964). DOI:10.1214/aoms/1177703732.

[9]. Ya. Z. Tsypkin, *Foundation of Information Identification Theory (in Russian).*(Nauka, Moscow, 1984)

[10]. H. Kodamana, B.Huang, R.Ranjan, Y.Zhao, R.Tan,N. Sammakneyad, Approaches to robust process identification. A review and tutorial of probabilistic methods. J. Process Control **66** (1),68-83 (2018).https://doi.org/10.1016/j.jprocont.2018.02.011.

[11]. F.R Hampel, The influence curve and its role in robust estimation, J. Am. Statist. Assoc. **69** (4), 383-393 (1974).DOI:10.1080/01621459.1974.10482962.

[12]. W.K. Ho, H.D.Ling,X. Vu, X. Wang, Filtering of the ARMAX process with generalized t – distribution noise: The influence function approach, Industrial & Engineering Chemical Research **53** (12),7019-7028 (2014). dx.doi.org/10.1021/ie401990x|Ind.Eng.Chem. Res.2014,53,7019-7028.

[13]. W.K. Ho, T. Chen, K.W.Ling, L. Sun L, Variance analysis of robust state estimation in power system using influence function. Electrical Power and Energy Systems**92** (10),53-63 (2017).http://dx.doi.org/10.1016/j.ijepes.2017.04.009.

[14]. N.N.R. Suri,N. Murty, G. Athithan,*Outliers Detection: Techniques and Application. A Data Mining Perspective.*(Springer, Berlin, 2019)

[15]. C.C. Aggarwal,*Outliers Analysis.*(Springer, Berlin , 2017)

[16]. M. Sugiyama, *Statistical Reinforcement Learning.*(CRC Press, Boca Raton, 2015)

[17]. M. Sugiyama,*Introduction to Statistical Machine Learning.*(Morgan Kaufman, London, 2016)

[18]. A.M. Zoubir, V. Koivunen, E. Ollila E, Muma,*Robust Statistics for Signal Processing.*(Cambridge University Press, Cambridge, 2018)

[19]. A. Sadeghian, O.Wu, B.Huang, Robust probabilistic principal component analysis based process modeling: Dealing with simultaneous contamination of both input and output data,J. Process Control **67** (3),94-111 (2018). https://doi.org/10.1016/j.jprocont.2017.03.012.

[20]. V. Filipovic,Recursive identification of multivariable ARX models in the presence of a priori information: Robustness and

regularization, Signal Processing**116** (4),68-77 (2015). http://dx.doi.org/10.1016/j.sigpro.2015.04.016.

[21]. V. Filipovic,Outlier robust stochastic approximation algorithm for identification of MIMO Hammerstein models, Nonlinear Dynamics **90** (8),1427-1441 (2017). http://DOI10.1007/s11071-017-3736-2.

[22]. V. Filipovic,A global convergent outlier robust adaptive predictor for MIMO Hammerstein models. International Journal of Robust and Nonlinear Control **27** (12),3350-3371 (2017).https://doi.org/10.1002/rnc.3705.

[23]. V. Filipovic, Recursive identification of block-oriented nonlinear systems in the presence of outliers, J. of Process Control**78** (6), 1-12 (2019). https://doi.org/10.1016/j.jprocont.2019.03.015.

[24]. J. Castro, A CTA model based on the Huber function, in: J. Domingo – Ferrer (Ed.),*Privacy in Statistical Data Bases*, (Springer, Berlin, 2014)

[25]. R. Holtey,*Multiple View Geometry in Computer Vision.*(Cambridge University Press, Cambridge, 2004)

[26]. L. Stefanski, D. Boss,The calculus of M – estimation,The American Statistician **56** (1): 29-38 (2002). DOI:10.1198/000313002753631330.

[27]. N. Stout,*Almost Sure Convergence.*(Academic Press, New York, 1974)

[28]. C. Desoer,M. Vidyasagar,*Feedback Systems: Input – Output Properties.(*Academic Pres,. New York, 1975)

[29]. V. Filipovic, Consistency of the robust recursive Hammerstein model identification algorithms, J. Frenkl. Inst.**352** (10),1932-1945 (2015). https://doi:10.1016/j..franklin.201502.005.

[30]. G.C. Goodwin,R.L. Payne,*Dynamic System Identification: Experiment Design and Data Analysis.*(Academic Press, New York, 1977)

[31]. J. Tukey, A survey for sampling from contaminated distributions. In: I. Olkin, (Ed.),*Contributions to Probability and Statistics.*(Stanford University Press, Stanford , 1960, pp. 448-485)