FACTA UNIVERSITATIS
Series: **Automatic Control and Robotics** Vol. 20, Nº 2, 2021, pp. 87 - 98
https://doi.org/10.22190/FUACR210321007F

Regular Paper

SYSTEM IDENTIFICATION USING NEWTON–RAPHSON METHOD BASED ON SYNERGY OF HUBER AND PSEUDO–HUBER FUNCTIONS

UDC (681.325:615.07)

Vojislav Filipović

University of Kragujevac, Faculty of Mechanical and Civil Engineering, Department of Automatic Control, Robotics and Fluid Technique, Kraljevo, Republic of Serbia

Abstract. *In real situations the presence of outliers is unavoidable and that is why the distribution of a disturbance is non-Gaussian. A synthesis of an algorithm of identification based on the Newton-Raphson method is considered for this case. The method requires that the loss function should be twice differentiable. Huber loss function, relevant for the treatment of outliers, has just the first derivative. In order to overcome the problem, the pseudo- Huber loss function is introduced. This function behaves similarly to the Huber loss function and has derivatives of an arbitrary order. In this paper, the pseudo- Huber loss function is used for the second derivative of functional in the Newton-Raphson procedure. The main contributions of the paper are: (i) Design of a new robust recursive algorithm based on the synergy of Huber and pseudo – Huber functions; (ii) The convergence analysis.*

Key words: *Robust identification, Huber function, Pseudo – Huber function, convergence analysis*

1. INTRODUCTION

Identification of a system is a very developed scientific field. There are numerous theoretical results [1-3]. One class of problems is dealing with robustness in the statistical sense including very low sensitivity to changes of the probability distribution of disturbance. For such class of problems the main tool is robust statistics [4]. The methodology is actual in various areas [5-7].

Received March 21, 2021 / Accepted June 09, 2021
Corresponding author: Vojislav Filipović
University of Kragujevac, Faculty of Mechanical and Civil Engineering, Department of Automatic Control, Robotics and Fluid Technique, Dositejeva 19, 36000 Kraljevo, Republic of Serbia
E-mail: v.filipovic@mts.rs

Practical researches have shown that the outliers are present in a real disturbance [8,9]. That is why the distribution of a disturbance is non-Gaussian. This fact is reflected in selection of the criteria of identification [10], which has a direct impact on the algorithm of identification.

The well known approach is based on the Huber loss function depending on the most unfavorable probability density of a disturbance. This function has only a derivative of the first order and the application of the Newton-Raphson method is not possible. In references [11] the problem is overcome by approximation of the relevant Fisher information. A smooth version of the Huber loss function is introduced in this paper (pseudo - Huber loss function) which has derivatives of an arbitrary order [12-14].

The paper proposes the Newton- Raphson algorithm in which the Huber loss function is used for the first derivative of functional, while for the second derivative of functional pseudo – Huber loss function is used. In the obtained algorithm, the gain matrix explicitly depends on the second derivative of the pseudo-Huber loss function.

The convergence analysis with probability one, based on the martingale theory [15], was performed for the proposed algorithm. Convergence problems are directly related to the presence of the pseudo - Huber function in the matrix gain of an algorithm. Conditions of persistent excitation depend on the conditional mathematical expectation of the matrix gain trace. For the Gaussian distribution they degenerate to standard conditions [16]. Also, the generalized strictly positive conditions are introduced through passive operators.

The main contributions of the paper are:

(i) The new robust recursive identification algorithm is proposed based on the Newton – Raphson algorithm, Huber`s loss function and pseudo - Huber loss function.

(ii) The convergence analysis of algorithm

2. ARMAX MODEL

The ARMAX model has a form

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k) \tag{1}$$

where $u(k) \in R^1$, $y(k) \in R^1$ and $e(k) \in R^1$ are input, output and stochastic disturbance respectively. Polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ are polynomials in the shift operator $q^{-1}y(k) = y(k-1)$ with

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + ... + a_n q^{-n} \\ B(q^{-1}) &= b_1 q^{-1} + ... + b_m q^{-m} \\ C(q^{-1}) &= 1 + c_1 q^{-1} + ... + c_r q^{-r} \end{aligned} \tag{2}$$

A common assumption is that the probability distribution of the stochastic disturbance $e(k)$ is known exactly. In what follows we will introduce the more realistic assumption about a class of distributions to which the disturbance belongs. The form of a class of distributions is

$$P_\varepsilon = \{P : P = (1-\varepsilon^*)N_D + \varepsilon^* G,\ G \text{ is symmetric}\} \tag{3}$$

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

$$N_D(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \tag{4}$$

The contamination model (3) for probability densities has a form

$$P_\varepsilon = \{p : p = (1-\varepsilon^*) p(e)(0, \sigma_N^2)_D + \varepsilon^* g(e)\} \tag{5}$$

where

$$p(e) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left\{-\frac{e^2}{2\sigma_N^2}\right\} \tag{6}$$

and $g(e)$ is a symmetric function.

Vector form of model (1) is

$$y(k) = \boldsymbol{\varphi}_0^T(k)\boldsymbol{\theta} + e(k) \tag{7}$$

where

$$\boldsymbol{\theta}^T = [a_1, ..., a_n, b_1, ..., b_m, c_1, ...c_r]$$

$$\boldsymbol{\varphi}_0^T(k) = \left[-y(k-1), .., -y(k-n), u(k-1), .., u(k-m), \mathrm{e}(k-1), .., \mathrm{e}(k-r)\right]$$

In the equation (7) $\boldsymbol{\varphi}_0(k)$ depends from the immeasurable quantity $e(i)$ ($i = k-1, k-2, \ldots, k-r$). The standard procedure in identification is to replace $e(k)$ with an estimated prediction error. We have

$$\boldsymbol{\varphi}^T(k) = \left[-y(k-1), .., -y(k-n), u(k-1), .., u(k-m), \varepsilon(k-1), .., \varepsilon(k-r)\right] \tag{8}$$

where

$$\varepsilon(k) = y(k) - \boldsymbol{\varphi}^T(k)\hat{\boldsymbol{\theta}}(k-1) \tag{9}$$

In relation (9) $\hat{\boldsymbol{\theta}}(k)$ is the estimate of the parameter $\boldsymbol{\theta}$.

3. Newton–Raphson Algorithm

Applying the Huber methodology [4], the least favourable probability density for a class of probability densities (5), is

$$p^*(e(k)) = \begin{cases} \dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_N} \exp\left\{-\dfrac{e^2(k)}{2\sigma_N^2}\right\}, & |e(k)| \le k_\varepsilon \\ \dfrac{1-\varepsilon}{\sqrt{2\pi}\sigma_N} \exp\left\{-\dfrac{k_\varepsilon}{\sigma_N^2}\left(|e(k)| - \dfrac{k_\varepsilon}{2}\right)\right\}, & |e(k)| > k_\varepsilon \end{cases} \tag{10}$$

where the relationship between the contamination degree ε and parameter k_ε of the Huber function is given with the next relation

$$\frac{2\Phi_N(k_\varepsilon)}{k_\varepsilon} - 2\Phi_N(-k_\varepsilon) = \frac{\varepsilon}{\varepsilon - 1}\ ,\ \Phi_N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \tag{11}$$

The last equation depends on the variables ε and k_ε. In practice, the contamination degree is unknown. Earlier intensive simulations [17] show that good performance of robust algorithms is provided for $k_\varepsilon \in [2, 4]$.

The Huber loss function is

$$\Phi(\varepsilon) = -\log p^*(e)\big|e = \varepsilon \tag{12}$$

or explicitly

$$\Phi(\varepsilon) = \begin{cases} \dfrac{\varepsilon^2(k)}{2\sigma_N^2} + \log\dfrac{\sqrt{2\pi}\sigma_N}{1-\varepsilon} & , \quad |\varepsilon| \le k_\varepsilon \\ \dfrac{k_\varepsilon}{\sigma_N^2}\left(|\varepsilon| - \dfrac{k_\varepsilon}{2}\right) + \log\dfrac{\sqrt{2\pi}\sigma_N}{1-\varepsilon} & , \quad |\varepsilon| > k_\varepsilon \end{cases} \tag{13}$$

Derivative of $\Phi(\cdot)$ is a Huber function

$$\psi(\varepsilon) = \begin{cases} \dfrac{\varepsilon}{\sigma_N^2} & , \quad |\varepsilon| \le k_\varepsilon \\ \dfrac{k_\varepsilon}{\sigma_N^2} sign(\varepsilon) & , \quad |\varepsilon| > k_\varepsilon \end{cases} \tag{14}$$

Usually in practice we take $\sigma_N^2 = 1$.

As one can see from relation (14) the Huber function is not differentiable in the two points ((k_ε) and ($-k_\varepsilon$)) . Owing to that fact the Huber loss function (13) is only first – order differentiable and it follows that is not applicable to second order method (for example Newton – Raphson algorithm, which is considered in this paper). Because we consider a smooth version of the Huber`s loss functions, the pseudo - Huber loss function which has derivatives of all degrees. [12-14].

In our case pseudo - Huber loss function has a form

$$\Phi_p(\varepsilon) = k_\varepsilon\left(\sqrt{(k_\varepsilon)^2 + \varepsilon^2} - k_\varepsilon\right) + \log\frac{\sqrt{2\pi}\sigma}{1-\varepsilon} \tag{15}$$

The functions $\Phi(\cdot)$ and $\Phi_p(\cdot)$ are close [13].

The derivatives of the loss function $\Phi_p(\cdot)$ (first and second) are

$$\psi_p(\varepsilon) = \Phi'_p(\varepsilon) = \frac{k_\varepsilon \varepsilon}{\sqrt{(k_\varepsilon)^2 + \varepsilon^2}} \tag{16}$$

$$\psi'_p(\varepsilon) = \Phi''_p(\varepsilon) = \frac{(k_\varepsilon)^3}{\left(\sqrt{((k_\varepsilon)^2 + \varepsilon^2)}\right)^{3/2}} \tag{17}$$

The derivatives of $\psi_p(\varepsilon)$ (that is pseudo - Huber function) and $\psi'_p(\varepsilon)$ are bounded and Lipschitz continuous [13].

Let us introduce next identification criteria

$$J(\boldsymbol{\theta}) = E\{\Phi(\varepsilon(k))\} \tag{18}$$

$$J^p(\boldsymbol{\theta}) = E\{\Phi_p(\varepsilon(k))\} \tag{19}$$

The corresponding empirical functionals are

$$J_k(\boldsymbol{\theta}) = \frac{1}{k}\sum_{i=1}^{k}\Phi(\varepsilon(k)) \tag{20}$$

$$J_k^p(\boldsymbol{\theta}) = \frac{1}{k}\sum_{i=1}^{k}\Phi_p(\varepsilon(k)) \tag{21}$$

In this paper we consider the following form of the modified Newton – Raphson algorithm

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) - [\nabla_\theta^2 J_k^p(\boldsymbol{\theta})]^{-1}[\nabla_\theta J_k(\boldsymbol{\theta})] = \hat{\boldsymbol{\theta}}(k-1) - [k\nabla_\theta^2 J_k^p(\boldsymbol{\theta})]^{-1}[k\nabla_\theta J_k(\boldsymbol{\theta})] \tag{22}$$

As in reference [11] we have

$$k\nabla_\theta J_k(\hat{\boldsymbol{\theta}}(k-1)) = -\psi(\varepsilon(k))\boldsymbol{\varphi}(k) \tag{23}$$

$$k\nabla_\theta^2 J_k^p(\hat{\boldsymbol{\theta}}(k-1)) = \sum_{i=0}^{k}\psi'_p(\varepsilon(i))\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^T(i) \tag{24}$$

Let us introduce

$$\boldsymbol{P}(k) = \left[\sum_{i=1}^{k}\psi'_p(\varepsilon(i))\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^T(i)\right]^{-1} \tag{25}$$

Applying the matrix inversion lemma from relations (22) – (25) we have recursive identification algorithm

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \boldsymbol{P}(k)\boldsymbol{\varphi}(k)\psi(\varepsilon(k)) \tag{26}$$

$$\boldsymbol{P}(k) = \boldsymbol{P}(k-1) - \frac{\boldsymbol{P}(k-1)\boldsymbol{\varphi}(k)\boldsymbol{\varphi}^T(k)\boldsymbol{P}(k-1)}{[\psi'_p(\varepsilon(k))]^{-1} + \boldsymbol{\varphi}^T(k)\boldsymbol{P}(k-1)\boldsymbol{\varphi}(k)} \tag{27}$$

with corresponding initial conditions.

Remark 1. Using approximation

$$\psi_a(\varepsilon(k)) \cong \psi_p(\varepsilon(k))$$

where

$$\psi_a(\varepsilon(k)) = \begin{cases} 1 & , \quad |\varepsilon(k)| \le k_\varepsilon \\ 0 & , \quad |\varepsilon(k)| > k_\varepsilon \end{cases}$$

It is possible to get the algorithms as in [10] and [17]. This approximation has a small influence on the behavior of gain of recursive algorithm.

In what follows we study the convergence of algorithms (26) – (27).

4. CONVERGENCE ANALYSIS

The convergence property of the algorithm (26)-(27) will be investigated using the martingale theory [15]. Throughout the following text we shall assume that $\{e(k)\}$ 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 stochastic disturbance

(A1) $\{e(k)\}$ is a sequence of independent and identically distributed random variables with symmetric distribution

(A2) All zeroes of the polynomial $C(q^{-1})$ are outside the unit circle

B) Hypotheses for the nonlinear function $\psi(\cdot)$

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

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

C) Hypotheses for the pseudo-Huber function

(C1) The function $\psi'_p(\cdot) \in [0,\infty)$

(C2) $$E\{\psi'_p(\varepsilon(k))\big|F_{k-1}\} = \Phi_{\psi p}\left(\frac{\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)$$

$$\Phi_{\psi p}(\cdot) \in (0,\infty), \quad \tilde{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k) - \boldsymbol{\theta}$$

D) Hypotheses for the conditional mathematical exception for trace of matrix gain $\boldsymbol{P}(k)$

$$r^a(k) = E\{tr\boldsymbol{P}^{-1}(k)\big|F_{k-1}\}$$

(D1) $$r^a(k) = r^a(k) + \Phi_{\psi p}\left(\frac{\tilde{\boldsymbol{\theta}}(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)\|\varphi(k)\|^2$$

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

E) Hypotheses for generalized strictly positive real conditions

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

$$Hz(k) = \Phi_1\left(\frac{z(k)}{C(q^{-1})}\right) - \frac{1}{2}\Phi_{\psi p}\left(\frac{z(k)}{C(q^{-1})}\right)z(k)$$

where

$$\Phi_1\left(\frac{z(k)}{C(q^{-1})}\right) = E\{\psi(\varepsilon(k))\big|F_{k-1}\},$$

$$z(k) = \tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)$$

F) Hypotheses about persistent excitation condition

(F1) There exists a constant $c > 0$ such that

$$\lim_{k\to\infty}\frac{\log^c(k_{\psi p}r^a(k))}{\lambda_{\min}\{\boldsymbol{P}(k)^{-1}\}} = 0, \; c > 1, \; k_{\psi p} > 1$$

The presented conditions (A-F) cover a large class of probability distributions (different choice of g(e) function). A special case is the Gaussian distribution.

Remark 2. The persistent excitation (condition F1) is

$$\lim_{k\to\infty}\frac{\log^c E\{(k_{\psi p})tr\boldsymbol{P}^{-1}(k)\big|F_{k-1}\}}{\lambda_{\min}\{\boldsymbol{P}(k)^{-1}\}}=0 \quad w.p.1$$

When probability distribution of stochastic disturbance is Gaussian then $\Phi_{\psi p}(\cdot)=1$, $k_{\psi p}=1$ and

$$E\{tr\boldsymbol{P}^{-1}(k)\big|F_{k-1}\}=E\left\{\sum_{i=1}^{k}\left\|\boldsymbol{\varphi}(i)\right\|^2\big|F_{k-1}\right\}=\sum_{i=1}^{k}\left\|\boldsymbol{\varphi}(i)\right\|^2=r_L(k)$$

Now condition (F1) has a form

$$\lim_{k\to\infty}\frac{\log^c r_L(k)}{\lambda_{\min}\{\boldsymbol{P}^{-1}(k)\}}=0 \quad w.p.1$$

and that is standard condition for linear algorithms [16].

Remark 3. The condition (E1) is based on the theory of passive operators [19].

Now we prove the following lemma.

Lemma: Consider the model (7) – (9) and algorithm (26) – (27) subject to the assumption (C1), (C2) and (D1). Then

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

Proof: Let us define the matrix

$$\begin{bmatrix}1 & \boldsymbol{\varphi}^T(k)\\ \boldsymbol{\varphi}(k) & \boldsymbol{P}^{-1}(k)\end{bmatrix} \tag{28}$$

Then Schur`s formula gives

$$\boldsymbol{\varphi}^T(k)\boldsymbol{P}(k)\boldsymbol{\varphi}(k)=\frac{\left|\boldsymbol{P}^{-1}(k)\right|-\left|\boldsymbol{P}^{-1}(k-1)\right|}{\left|\boldsymbol{P}^{-1}(k)\right|} \tag{29}$$

where $|\cdot|$ denotes the determinant.

Let us notice

$$\left|\boldsymbol{P}^{-1}(k)\right|=\prod_{k=1}^{d}\lambda_k\left\{\boldsymbol{P}^{-1}(k)\right\}\le\lambda_{\max}^{d}\left\{\boldsymbol{P}^{-1}(k)\right\}\ , \tag{30}$$

$$d=n+m+r$$

Now introduce

$$r(k)=tr\{P^{-1}(k)\}=\sum_{i=1}^{k}\psi_p'(\varepsilon(i))\left\|\boldsymbol{\varphi}(i)\right\|^2 \tag{31}$$

For the prediction error is

$$\varepsilon(k)=-\frac{\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}+e(k) \tag{32}$$

Using relation (31) and condition (D1) and (C2) we have

$$E\{r(k)\big|F_{k-1}\} = E\left\{\sum_{i=1}^{k}\psi'_p(\varepsilon(i))\|\boldsymbol{\varphi}(i)\|^2\,\big|F_{k-1}\right\} = \sum_{i=1}^{k}\Phi_{\psi p}\left(\frac{z(k)}{C(q^{-1})}\right)\|\boldsymbol{\varphi}(i)\|^2 = r^a(k) \quad (33)$$

From assumption (D1) it follows that

$$\liminf_{k\to\infty}\{r(k)\big|F_{k\text{-}1} =\}¥ \quad (34)$$

and consequently

$$\liminf_{k\to\infty} r(k) = \infty, \quad w.p.1 \quad (35)$$

Using conditions (C1) and (C2) we have

$$k_{\psi p}\Phi_{\psi p}\left(\frac{\tilde{\boldsymbol{\theta}}(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right) \geq \psi'_p(\varepsilon(k)) \quad (36)$$

From relation (36) it follows that

$$r(k) \leq k_{\psi p}\sum_{i=1}^{k}\Phi_{\psi p}\left(\frac{\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)\|\boldsymbol{\varphi}(i)\|^2 = k_{\psi p}r^a(k) \quad (37)$$

For matrix gain and sequence $r(k)$ is valid

$$\left|\boldsymbol{P}^{-1}(k)\right| = \prod_{k=1}^{d}\lambda_k\{\boldsymbol{P}^{-1}(k)\} \leq \lambda^d_{\max}\{\boldsymbol{P}^{-1}(k)\} \quad (38)$$

$$r(k) = \sum_{k=1}^{d}\lambda_k\{\boldsymbol{P}^{-1}(k)\} \geq \lambda_{\max}\{\boldsymbol{P}^{-1}(k)\} \quad (39)$$

Using relation (37) we have

$$\sum_{k=k_0}^{\infty}\frac{\varphi^T(k)P(k)\varphi(k)}{\log^c(k_{\psi p}r^a(k))} \leq \sum_{k=k_0}^{\infty}\frac{\varphi^T(k)P(k)\varphi(k)}{\log^c r(k)} \quad (40)$$

From (38) and (39) it follows

$$r(k) \geq \left|\boldsymbol{P}^{-1}(k)\right|^{1/d} \quad (41)$$

Using relations (29), (35), (40) and (41) we have

$$\begin{aligned}\sum_{k=k_0}^{\infty}\frac{\boldsymbol{\varphi}^T(k)\boldsymbol{P}(k)\boldsymbol{\varphi}(k)}{\log^c r(k)} &\leq d^c\sum_{k=k_0}^{\infty}\frac{\left|\boldsymbol{P}^{-1}(k)\right| - \left|\boldsymbol{P}^{-1}(k-1)\right|}{\left|\boldsymbol{P}^{-1}(k)\right|\log^c\left|\boldsymbol{P}^{-1}(k)\right|} \leq \\ \leq d^c\sum_{k=k_0}^{\infty}\int_{\left|\boldsymbol{P}^{-1}(k_0)\right|}^{\left|\boldsymbol{P}^{-1}(k)\right|}\frac{dt}{\log^c t} &= \frac{d^c}{(c-1)\log^c\left|\boldsymbol{P}^{-1}(k_0)\right|} < \infty \ , w.p.1\end{aligned} \quad (42)$$

In the relation (42) two facts are used

$$\left|\boldsymbol{P}^{-1}(k)\right| > 1, \ \text{for} \ \forall k \geq k_0 \quad (43)$$

$$\left|\boldsymbol{P}^{-1}(\infty)\right| = \infty \ , \ w.p.1 \quad (44)$$

The last equation follows from relation (35). From relation (40) and (42) it follows the statement of Lemma. ▪

Now we shall formulate the main result. The proof of the theorem is similar to the proof in the reference [18], but is given for completeness.

Theorem: Let us suppose that for model (7) – (9) and algorithm (26) – (27) the assumptions of the Lemma are fulfilled and assume that the following hypotheses are satisfied: (A1) – (A2), (B1) – (B2), (E1) and (F1). Then

$$P\{\lim_{k\to\infty}\hat{\boldsymbol{\theta}}(k)=\boldsymbol{\theta}\}=1 \text{ ▪}$$

Proof: Let us introduce the stochastic Lyapunov`s function

$$V(k)=\tilde{\boldsymbol{\theta}}^T(k)\boldsymbol{P}^{-1}(k)\tilde{\boldsymbol{\theta}}(k) \tag{45}$$

Using (26) and (45) we have

$$V(k)=\tilde{\boldsymbol{\theta}}(k-1)\boldsymbol{P}(k)\tilde{\boldsymbol{\theta}}(k-1)+2\tilde{\boldsymbol{\theta}}(k-1)\boldsymbol{\varphi}(k)\psi(\varepsilon(k))+\boldsymbol{\varphi}^T(k)\boldsymbol{P}(k)\boldsymbol{\varphi}(k)\psi^2(\varepsilon(k)) \tag{46}$$

From (21) and (27) it follows

$$P^{-1}(k)=P^{-1}(k-1)+\psi'_p(\varepsilon(k))\varphi(k)\varphi^T(k) \tag{47}$$

Using (46) and (47) one obtains

$$\begin{aligned}V(k)=V(k-1)+2\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)\psi(\varepsilon(k))+\psi'_p(\varepsilon(k))(\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k))^2+\\+\boldsymbol{\varphi}^T(k)\boldsymbol{P}(k)\boldsymbol{\varphi}(k)\psi^2(\varepsilon(k))\end{aligned} \tag{48}$$

We now define next function (using relation (32)) and according with assumptions (A1) and (A2)

$$\Phi_2\left(\frac{\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)=E\{\psi^2(-\varepsilon(k))\big|F_{k-1}\} \tag{49}$$

Having in mind assumptions (B2) one concludes

$$\Phi_2\left(\frac{\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)\le k_1 \quad , \; k_1\in(0,\infty) \tag{50}$$

According with assumption (A1), (B1), (E1) and (50) we have

$$\begin{aligned}E\left\{\frac{V(k)}{\log^c(k_{\psi p}r^a(k))}\big|F_{k-1}\right\}\le\frac{V(k-1)}{\log^c(k_{\psi p}r^a(k))}-\frac{2\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{\log^c(k_{\psi p}r^a(k))}\left[\Phi_1\left(\frac{\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)-\right.\\\left.-\frac{1}{2}\Phi_{\psi p}\left(\frac{\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)\tilde{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)\right]+k_1\frac{\boldsymbol{\varphi}^T(k)\boldsymbol{P}(k)\boldsymbol{\varphi}(k)}{\log^c(k_{\psi p}r^a(k))}\end{aligned} \tag{51}$$

From (E1) it follows that

$$\mathrm{S}(k)=2\sum_{i=1}^{k}\tilde{\boldsymbol{\theta}}^{T}(i-1)\boldsymbol{\varphi}(i)\ \left[\Phi_1\left(\frac{\tilde{\boldsymbol{\theta}}^{T}(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)-\right.$$
$$\left.-\frac{1}{2}\Phi_{\psi p}\left(\frac{\tilde{\boldsymbol{\theta}}^{T}(k-1)\boldsymbol{\varphi}(k)}{C(q^{-1})}\right)\tilde{\boldsymbol{\theta}}^{T}(k-1)\boldsymbol{\varphi}(k)\right]+k_2,k_2\in(0,\infty) \quad (52)$$
$$k\to\infty$$

Let us define a quantity

$$T(k)=R(k)+\frac{S(k)}{\log^{c}(k_{\psi p}r^{a}(k))},\ R(k)=\frac{V(k)}{\log^{c}(k_{\psi p}r^{a}(k))} \quad (53)$$

From (51 – 53) it follows that

$$E\{T(k)\big|F_{k-1}\}\le T(k-1)+k_1\frac{\boldsymbol{\varphi}^{T}(k)\boldsymbol{P}(k)\boldsymbol{\varphi}(k)}{\log^{c}(k_{\psi p}r^{a}(k))} \quad (54)$$

From relation (25), Lemma and martingale convergence theorem we have

$$\lim_{k\to\infty}T(k)=T^{*} \quad (55)$$

From the last relation one obtains

$$\lim_{k\to\infty}R(k)=R^{*},\ \ w.p.1 \quad (56)$$

Further we have

$$R(k)=\frac{tr\{\boldsymbol{P}(k)^{-1}\boldsymbol{\varphi}(k)\boldsymbol{\varphi}^{T}(k)\}}{\log^{c}(k_{\psi p}r^{a}(k))}\le\frac{\lambda_{\min}\{\boldsymbol{P}(k)^{-1}\}\left\|\tilde{\boldsymbol{\theta}}(k)\right\|^{2}}{\log^{c}(k_{\psi p}r^{a}(k))}=\frac{\left\|\tilde{\boldsymbol{\theta}}(k)\right\|^{2}}{\dfrac{\log^{c}(k_{\psi p}r^{a}(k))}{\lambda_{\min}\{\boldsymbol{P}(k)^{-1}\}}} \quad (57)$$

From assumption (F1) and relation (56), (57) follows the proof of theorem. ▪

5. SIMULATION STUDY

We will consider the next ARMAX model

$$A(q^{-1})y(k)=B(q^{-1})u(k)+C(q^{-1})e(k)$$

$$A(q^{-1})=1-0.85q^{-1}+0.6q^{-2}-0.7q^{-3}$$

$$B(q^{-1})=0.8q^{-1}-0.5q^{-2}$$

$$C(q^{-1})=1-0.4q^{-1}$$

It is supposed that the stochastic disturbance has a non-Gaussian distribution

$$e\sim(1-\varepsilon)^{*}N(0,\sigma_1^{2})+\varepsilon^{*}N(0,\sigma_2^{2})$$

where $N(0,\sigma^{2})$ is a Gaussian distribution with mean m and variance σ^{2}. It is supposed that

$$\sigma_1^{2}=1\ ,\ \sigma_2^{2}=100$$

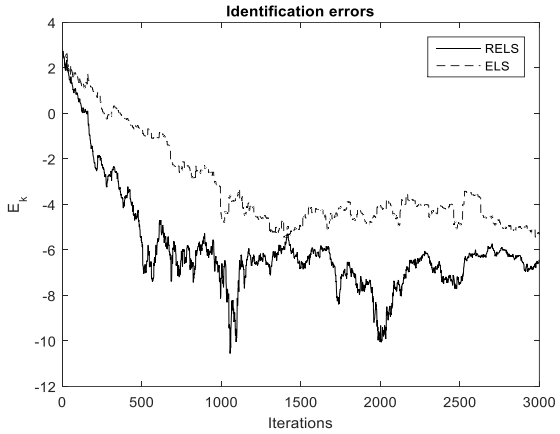

Fig. 1 Comparison of RELS and ELS for $\varepsilon^* = 0.1$

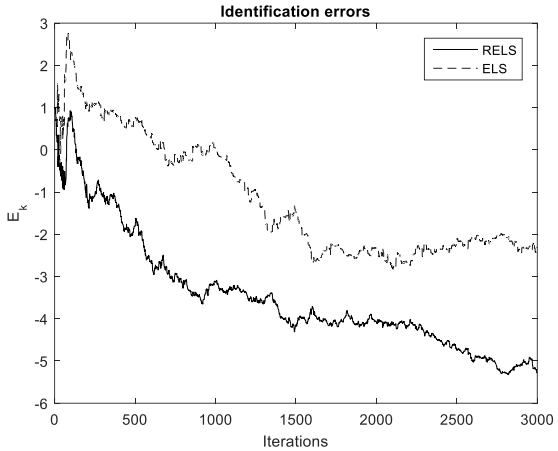

Fig. 2 Comparison of RELS and ELS for $\varepsilon^* = 0.2$

The form of the estimation error is

$$E_k = \ln\left\|\tilde{\theta}(k) - \theta\right\|^2.$$

We will consider the following types of errors

ELS (Extended least squares) - for standard linear algorithms $(\psi(x) = x)$

RELS (Robust extended least squares) – for robust algorithms (26) – (27)

It is considered to have the following degrees of contamination: $\varepsilon^* = 0.1; 0.2$. The Huber parameter is $k_\varepsilon = 3$.

From above figures it is possible to conclude that the algorithm proposed in the paper is superior in comparison with ELS (extended least squares).

6. CONCLUSION

Paper considers the Newton–Raphson algorithm for the case when observations have outliers. The method requests that the loss function, relevant for criterion identification, is second order differentiable. The Huber loss function has only the first derivative. The pseudo - Huber`s loss function has derivatives of all degrees and behaves similarly as the Huber loss function. The recursive algorithm is based on the synergy of both functions. The convergence analysis is performed. Further investigations will be related to identification of nonlinear and multivariable systems.

REFERENCES

[1] F. Huang, J. Zhang, S. Zhang, "Mean – square – deviation analysis of probabilistic LMS algorithm," Digit. Signal Process, vol.92, no.9, pp 26-35, 2019. doi.org/10.1016/j.dsp.2019.102582.
[2] F.S.L.G. Duarte, R.A. Rios, E.R. Hruschka, R.F. de Mello, "Decomposing time series into deterministic and stochastic influences: A survey," Digit. Signal Process, vol.95, no.12, 102582, 2019.doi.org/10.1016/j.dsp. 2019.102582.
[3] S.Y. Wang, W. Wang, L. Y. Dang, Y. X. Jiang, "Kernel least mean – square based on the Nystrom method," Circuit, Systems and Signal Processing, vol.38, no.11, pp.3133-3151,2019. doi: 10.1007/s00034-018-1006-2.
[4] P. Huber, E. Ronchetti, Robust Statistics, Wiley, New York, 2009.
[5] M. Sugiyama, Introduction to Statistical Machine Learning, Morgan Kaufman, New York, 2016.
[6] N.N.R. Suri, N. Murty, G. Athithan, Outlier Detection: Techniques and Application. A Data Mining Perspective, Springer, Berlin, 2019.
[7] A.M. Zoubir, V. Koivunen, E. Ollila, M. Muma, Robust Statistics for Signal Processing, Cambridge University Press, Cambridge, 2018.
[8] R.A. Maronna, R.D. Martin, V. Yohai, M. Salibian – Barrera, Robust Statistics. Theory and Methods (with R), Wiley, New York, 2019.
[9] R. Pearson, Exploring Data in Engineering, the Science and Medicine, Oxford University Press, Oxford, 2011.
[10] Ya. Z. Tsypkin, On the Foundations of Information and Identification Theory (in Russian), Nauka, Moscow, 1984.
[11] V. Filipovic, "Recursive identification of block-oriented nonlinear systems in the presence of outliers", J. of Process Control, vol.78, no. 6, pp. 1-12, 2019. doi: 10.1016/j.jprocont.2019.03.015.
[12] J. Castro, A CTA model based on the Huber function, J. Domingo – Ferrer (Ed.): Privacy in Statistical Data Bases, Springer, Berlin, 2014.
[13] R. Holtey, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2004.
[14] L. Stefanski, D. Boss, "The calculus of M – estimation. The American Statistician," vol.56, no. 1, pp.29-38,2002. doi: 10.1198/000313002753631330.
[15] N. Stout, Almost Sure Convergence, Academic Press, New York, 1974.
[16] L. Lai, Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," Annals of Statistics, vol.10, no.1, pp.151-166,1982. doi: 10.2307/2240506.
[17] V. Filipovic, "Consistency of the robust recursive Hammerstein model identification algorithm," J. Franklin Inst., vol.352 , no.5, pp. 1932-1943, 2015. doi: 101016//j.jfranklin.201502.005.
[18] V. Filipovic, B. Kovacevic, "On robust AML identification algorithms," Automatica, vol.30, no.11, pp. 1775-1778,1994. doi: 10.1016/0005-1098(94) 90081-7.
[19] C. Desoer, M. Vidyasagar, Feedback Systems: Input - Output Properties, Academic Press, New York, 1975.