

## ROBUST RECURSIVE IDENTIFICATION OF HAMMERSTEIN MODELS BASED ON WEISZFALD ALGORITHM

UDC (681.5.015:004.896)

**Vojislav Filipović**

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

**Abstract.** *The Hammerstein models can accurately describe a wide variety of nonlinear systems (chemical process, power electronics, electrical drives, sticky control valves). Algorithms of identification depend, among other, on the assumption about the nature of stochastic disturbance. Practical research shows that disturbances, owing the presence of outliers, have a non-Gaussian distribution. In such case it is a common practice to use the robust statistics. In the paper, by analysis of the least favourable probability density, it is shown that the robust (Huber's) estimation criterion can be presented as a sum of non-overlapping  $l_1$  - norm and  $l_2$  - norm criteria. By using a Weiszfeld algorithm  $l_1$  - norm criterion is converted to  $l_2$  - norm criterion. So, the weighted  $l_2$  - norm criterion is obtained for the identification. The main contributions of the paper are: (i) Presentation of the Huber's criterion as a sum of  $l_1$  - norm and  $l_2$  - norm criteria; (ii) Using the Weiszfeld algorithm  $l_1 - l_2$  - norm criterion is converted to a weighted  $l_2$  - norm criterion; (iii) Weighted extended least squares in which robustness is included through weighting coefficients are derived for NARMAX (nonlinear autoregressive moving average with exogenous variable) . The illustration of the behaviour of the proposed algorithm is presented through simulations.*

**Key words:** *Hammerstein model, non-Gaussian noise, Weiszfeld algorithm, weighted least squares*

### 1. INTRODUCTION

The main components of system identification are: the observed data, a set of candidate models, a criterion fit and validation [1]. In that context it is important how to model disturbance. In the stochastic frame it is supposed that the probability distribution of disturbance is known. In most cases it is supposed that disturbance has a Gaussian

---

Received September 24, 2019

**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

distribution. Practice showed that it is not correct owing to [?] the outliers presence [2]. Such fact causes that the standard algorithm significantly reduced efficiency. The imperative is to design an algorithm which has a low sensitivity to change in disturbances distribution. The tool for the design of such an algorithm is the robust statistics [3].

The actual research in the field of identification devotes considerable attention to robust identification methods [4]. The robust identification of multi-input multi-output models using the stochastic approximation is considered in [5] and the robust adaptive prediction is presented in [6]. The key ingredient in the robust statistics is the Huber loss function. But that function is only the first order differentiable. It means that the second order methods (Newton – Raphson methodology) cannot be used. The problem can be avoided by the approximation of the Fisher information for the least favourable probability density of the stochastic disturbance [7]. The field of machine learning also includes robust estimation procedures [8].

In this paper we consider a design of a robust recursive identification algorithm. The problem belongs to the robust identification of block-oriented nonlinear models [9,10]. The nonlinear model has a structure of a NARMAX model. It means that the model consists of a static nonlinear element coupled in series with a linear time – invariant dynamic system. Such a model is known as the Hammerstein model. The static nonlinearity is described by a polynomial function of the input while the linear part is described by an ARMAX model. In this paper, we assume that observations include outliers and as a disturbance model a class of distributions (mixture of Gaussian and arbitrary symmetric distribution) is used .

In this paper, it is started from the fact that the criterion of identification, based on the robust statistics (least favourable probability density), is a  $l_1 - l_2$  – norm estimation problem. Using the Weiszfeld algorithm [11,12],  $l_1$  part of the above criterion is converted to the  $l_2$ - norm. From that it follows that the identification criterion can be presented in the form of a generalized  $l_2$ - norm estimation problem. The application of the Weiszfeld algorithm introduces a level of non-optimality in a resultant  $l_2$  - norm criterion [13]. But the benefits are the significant reduction of [required] computation. By using the novel  $l_2$  - norm criterion, a robust recursive algorithm for identification of the NARMAX model. is derived. The main contributions of the paper are: (i) It is showed that the Huber`s criterion is a sum of non – overlapping  $l_1$  - norm and  $l_2$  - norm criterions; (ii) By using the Wieszfald algorithm the  $l_1 - l_2$  – norm criterion is converted to the weighted  $l_2$  - norm criterion; (iii) The design of the novel robust extended least squares algorithm; (iv) For derivation of the recursive algorithm the second derivative of Huber`s loss function which does not exist is not necessary.

The paper is organized as follows. Section 2 describes the history of the Wieszfald algorithm. In Section 3 a new form of identification criterion is introduced. Section 4 presents the novel recursive algorithm for NARMAX models. The illustration of the algorithm behaviour is given in the Section 5 by simulation and concluding remarks are given in the last section.

## 2. HISTORY

The location problem is very important and is the subject of intense research [14,15]. In what follows will be described the history of the location problem.

At the beginning of the seventeenth century, French mathematician Fermat formulated the next problem [16]

*Given three points in a plane, find a fourth point such that the sum of its distances to the three given points is as small as possible.*

The Italian physicist and mathematician Torricelli found a geometrical method for solving the above problem. The extension to four given points in the plane is easy and was made by Fagnano (1775). The move to more than four given points in the plane was made by known mathematicians such as Gauss, Steiner and others.

At the beginning of the twentieth century, the German economist Alfred Weber made a generalization of the Fermat problem by assigning different weights to the known points [17]. So the Fermat problem is transformed into an industrial problem in which a plant is to be located (the unknown point) so as to minimize transportation costs from suppliers to consumers (the known points) requiring different amounts of products (the weights). Practically, the location of a plant should be chosen so that the transportation costs of raw materials and products are the lowest possible. In the literature the location problem is often called the Fermat–Weber problem.

The location problem in the computer science (more specifically analysis in computational geometry) is determined

*The minimum number of equal geometrical shapes that are required to cover a certain area and the positions of their centroids.*

Generally we can formulate the location problem in a more abstract manner [14]

*Given some metric space and a set of known points, determine a number of additional points so as to optimize a function of the distance between new and existing points.*

In what follows we describe the main idea of the Weiszfeld algorithm. Let us suppose that the process is described with the regression model

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

where  $y(k) \in R^1$  is the output of the system,  $e(k) \in R^1$  is a stochastic disturbance,  $\boldsymbol{\varphi}(\cdot) \in R^d$  is a vector of measurement and  $\boldsymbol{\theta} \in R^d$  is a vector of parameters.

The least–modules method can be given by minimization of the criterion [18]

$$J_L(\boldsymbol{\theta}) = \sum_{i=1}^N |y(i) - \boldsymbol{\varphi}^T(i)\boldsymbol{\theta}| \rightarrow \inf_{\boldsymbol{\theta} \in R^d} \quad (2)$$

The criterion (2) generates the procedure which is outlier robust.

Minimization of the criterion (2) can be performed by using the linear programming and the method of internal point [19,20]. But, in that case the problem becomes numerically very complex.

The Weiszfeld method is attractive for solving the problem (2). Weiszfeld developed an algorithm that solved the Weber problem with an arbitrary number of customers [11,12].

The idea of Weiszfeld algorithms lies in performing minimization of non smooth functional  $J_L$  by using the iterative procedure. It is considered functional

$$J_W(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k-1)}) = \sum_{i=1}^N w(i)^{(k-1)} |y(i) - \boldsymbol{\varphi}^T(i)\boldsymbol{\theta}|^2 \quad (3)$$

where  $k$  is the number of iterations, weight coefficient

$$w(i)^{(k-1)} = \frac{1}{|y(i) - \boldsymbol{\varphi}^T(i)\boldsymbol{\theta}^{(k-1)}|} \quad (4)$$

And  $\boldsymbol{\theta}^{(k-1)}$  is the vector of parameters obtained of the preceding iteration. A sequence of vector  $\boldsymbol{\theta}^{(k)}$  is given from the next minimization problem

$$\boldsymbol{\theta}^{(k)} = \arg \min_{\boldsymbol{\theta}} J_w(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k-1)}) \quad (5)$$

Finally we remark that the Weiszfeld method remained unknown until the Kuhn seminar given in Budapest 1963 [21].

### 3. GENERALIZED IDENTIFICATION CRITERION

In this part of the paper we will define different norms relevant for determination of the identification criterion. Let us suppose that  $\{f_k\}$  is the function sequence. A  $p$ -norm is given as [22]

$$\|f_p\| = \left[ \sum_{k=1}^n |f_k|^p \right]^{\frac{1}{p}}, \quad 0 \leq p < \infty \quad (6)$$

Relevant norms for this paper are

$$\|f\|_1 = \sum_{k=1}^n |f_k|, \quad l_1\text{-norm} \quad (7)$$

$$\|f\|_2 = \left[ \sum_{k=1}^n |f_k|^2 \right]^{\frac{1}{2}}, \quad l_2\text{-norm} \quad (8)$$

We will suppose that the stochastic disturbance  $e(k)$ , in model (1), has a non-Gaussian distribution with next structure

$$P_\varepsilon = \{P : P = (1 - \varepsilon)N(0, \sigma_N^2) + \varepsilon g(e)\} \quad (9)$$

where  $\varepsilon \in [0, 1)$  is the contamination degree,  $N(0, \sigma_N^2)$  is the Gaussian probability density

$$N(0, \sigma_N^2) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left\{-\frac{e^2}{2\sigma_N^2}\right\} \quad (10)$$

and  $g(e)$  is the arbitrary symmetric probability density.

According to the Huber's methodology the least favourable probability density, for the class of probability densities (9), is [3]

$$p^*(e(k)) = \begin{cases} \frac{1-\varepsilon}{\sqrt{2\pi}\sigma_N} \exp\left\{-\frac{e^2(k)}{2\sigma_N^2}\right\}, & |e(k)| \leq k_\varepsilon \\ \frac{1-\varepsilon}{\sqrt{2\pi}\sigma_N} \exp\left\{-\frac{k_\varepsilon}{\sigma_N^2} \left(|e(k)| - \frac{k_\varepsilon}{2}\right)\right\}, & |e(k)| > k_\varepsilon \end{cases} \quad (11)$$

where the relationship between the contamination degree  $\varepsilon$  and the parameter  $k_\varepsilon$  is

$$\frac{2\Phi_N(k_\varepsilon)}{k_\varepsilon} - 2N\Phi_N(k_\varepsilon) = \frac{\varepsilon}{\varepsilon-1}, \quad \Phi_N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \quad (12)$$

In relation (12) the contamination degree  $\varepsilon$  is unknown. It follows from simulations that good performances of the identification algorithm are provided for  $k_\varepsilon \in [2, 4]$  [7].

In the frame of robust statistics the criterion identification has a form.

$$\Phi(\varepsilon_p(k)) = -\log p^*(\varepsilon_p(k)) \quad (13)$$

From relation (11) and (13) it follows that

$$\Phi(\varepsilon_p(k)) = \begin{cases} \frac{\varepsilon_p(k)}{\sqrt{2\pi}\sigma_N} + \ln \frac{\sqrt{2\pi}\sigma_N}{1-\varepsilon}, & |\varepsilon_p(k)| \leq k_\varepsilon \\ \frac{k_\varepsilon}{\sigma_N^2} \left(|\varepsilon_p(k)| - \frac{k_\varepsilon}{2}\right) + \ln \frac{\sqrt{2\pi}\sigma_N}{1-\varepsilon}, & |\varepsilon_p(k)| > k_\varepsilon \end{cases} \quad (14)$$

The analysis of the loss function (14) shows that its form is equivalent to the form of the loss function for the Gaussian distribution for  $|\varepsilon_p(k)| \leq k_\varepsilon$  ( $l_2$  - loss function) and that is equivalent to the form of the loss function for the Laplace distribution for  $|\varepsilon_p(k)| > k_\varepsilon$  ( $l_1$  - loss function). Based on that, using the concept of empirical functional, equivalent criterion identification is

$$J(\theta) = \frac{1}{N} \sum_{|\varepsilon_p(k)| \leq k_\varepsilon} \alpha^1(k) \varepsilon_p^2(k) + \sum_{|\varepsilon_p(k)| > k_\varepsilon} k_\varepsilon |\varepsilon_p(k)| = J_G(\theta) + J_L(\theta) \quad (15)$$

The criterion (15) is  $l_1 - l_2$  - identification criterion

Let us consider the first term in the relation (15).

$$J_G(\theta) = \frac{1}{N} \sum_{|\varepsilon_p(k)| \leq k_\varepsilon} \alpha^1(k) \varepsilon_p^2(k) = \sum_{|\varepsilon_p(k)| \leq k_\varepsilon} \alpha^1(k) (y(k) - \theta^T \varphi(k))^2 \quad (16)$$

The coefficients  $\{\alpha^1(k)\}$  constitute a sequence of positive numbers. The coefficients  $\alpha^1(k)$  gives different weights to different observations [23]. Most often in algorithms  $\alpha^1(k)$  is chosen equal to one. The optimal weighting factor  $\alpha^1(k)$ , in the least squares criterion (16) is the inverse of the variance of the stochastic disturbance, i.e.

$$\alpha^1(k) = \frac{1}{\sigma_N^2} \quad (17)$$

Let us now consider the second term in the relation (15). According to the Weiszfeld algorithm we have

$$\begin{aligned} k_\varepsilon |y(k) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(k)| &= \frac{k_\varepsilon}{|y(k) - \hat{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)|} (y(k) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(k))^2 = \\ &= \alpha^2(k) (y(k) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(k))^2 \end{aligned} \quad (18)$$

From (18) it follows that

$$J_L(\boldsymbol{\theta}) = \frac{1}{N} \sum_{|\varepsilon_p(k)| > k_\varepsilon} \alpha^2(k) (y(k) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(k))^2 \quad (19)$$

Using relation (15), (16) and (19) we have

$$J(\boldsymbol{\theta}) = \frac{1}{N} \sum_{k=1}^N \alpha(k) (y(k) - \boldsymbol{\theta}^T \boldsymbol{\varphi}(k))^2 \quad (20)$$

$$\alpha(k) = \begin{cases} \alpha^1(k) & , \quad |\varepsilon_p(k)| \leq k_\varepsilon \\ \alpha^2(k) & , \quad |\varepsilon_p(k)| > k_\varepsilon \end{cases} \quad (21)$$

where

$$\alpha^1(k) = \frac{1}{\sigma_N^2} \quad (22)$$

$$\alpha^2(k) = \frac{k_\varepsilon}{|y(k) - \hat{\boldsymbol{\theta}}^T(k-1)\boldsymbol{\varphi}(k)|} \quad (23)$$

The relations (14)–(23) show to us that using the Weiszfeld algorithm  $l_1 - l_2$  – the identification criterion is converted to the weighted  $l_2$  – identification criterion. Property of robustness is contained in the coefficient  $\alpha(k)$ .

**Remark 1.** In reference [24] the next loss function is proposed

$$L(x, \alpha, c) = \frac{|2 - \alpha|}{\alpha} \left( \left( \frac{(x/c)^2}{|2 - \alpha|} + 1 \right)^{\frac{\alpha}{2}} - 1 \right) \quad (24)$$

In the relation (24)  $\alpha \in R^1$  is a shape parameter that controls the robustness of the loss function and  $c \in R^1$   $c > 0$  is a scale parameter that control the size of the quadratic loss near  $x = 0$ .

The loss function (24) is undefined when  $\alpha = 2$  but it approaches  $l_2$  - loss function (squared prediction errors) in the limit

$$\lim_{\alpha \rightarrow 2} L(x, \alpha, c) = \frac{1}{2} \left( \frac{x}{c} \right)^2 \quad (25)$$

When  $\alpha = 1$  the function (24) is a smoothed form of  $l_1$  - loss function

$$L(x, 1, c) = \sqrt{\left( \frac{x}{c} \right)^2 + 1} - 1 \quad (26)$$

The loss function (26) known as the pseudo – Huber function.

#### 4. RECURSIVE IDENTIFICATION OF NARMAX MODEL

In this section we consider identification of the NARMAX system using the identification criterion (20). We will consider a block-oriented nonlinear model that consists of the interaction of linear time – invariant dynamic subsystem and a static nonlinear element.

It is supposed that the nonlinear part of the nonlinear (Hammerstein) model is a polynomial of a known order in the input as follows

$$v(k) = f(u(k)) = d_1 \gamma_1(u(k)) + d_2 \gamma_2(u(k)) + \dots + d_s \gamma_s(u(k)) \quad (27)$$

with the known basis function  $\gamma_1, \gamma_2, \dots, \gamma_s$ . For the unique parameterization of the Hammerstein model it is necessary that  $d_1 = 1$  [7].

The linear part of the model is described as an ARMAX model

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k) \quad (28)$$

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

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) &= b_1 q^{-1} + \dots + b_m q^{-m} \\ C(q^{-1}) &= 1 + c_1 q^{-1} + \dots + c_r q^{-r} \end{aligned} \quad (29)$$

From (27) – (29) it follows that

$$\begin{aligned} y(k) &= -\sum_{i=1}^n a_i y(k-i) + \sum_{i=1}^m b_i u(k-i) + \sum_{i=1}^r c_i e(k-i) + e(k) = \\ &= -\sum_{i=1}^n a_i y(k-i) + \sum_{i=1}^m b_i \sum_{j=1}^s d_j \gamma_j(u(k-j)) + \sum_{i=1}^r c_i e(k-i) + e(k) \end{aligned} \quad (30)$$

Let us define the following vectors

$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{a} \\ d_1 \mathbf{b} \\ \vdots \\ d_s \mathbf{b} \\ \mathbf{c} \end{bmatrix} \in R^{n_0}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \end{bmatrix} \quad (31)$$

Where  $n_0 = n + sm + r_s$  and

$$\mathbf{h}_0(k) = \begin{bmatrix} -y(k-1) \\ -y(k-2) \\ \vdots \\ -y(k-n) \end{bmatrix}, \quad \mathbf{h}_j(k) = \begin{bmatrix} \gamma_j(u(k-1)) \\ \gamma_j(u(k-2)) \\ \vdots \\ \gamma_j(u(k-m)) \end{bmatrix}, \quad \mathbf{h}(k) = \begin{bmatrix} h_0(k) \\ h_1(k) \\ \vdots \\ h_s(k) \end{bmatrix}, \quad \boldsymbol{\varphi}_0(k) = \begin{bmatrix} \mathbf{h}(k) \\ e(k-1) \\ e(k-2) \\ \vdots \\ e(k-r) \end{bmatrix} \quad (32)$$

where  $j = 1, 2, \dots, s$  and  $\mathbf{h}_0(k) \in R^n$ ,  $\mathbf{h}_j(k) \in R^m$ ,  $\mathbf{h}(k) \in R^{n+sm}$  and  $\boldsymbol{\varphi}(k) \in R^{n_0}$ .

From (30) – (32) it follows that

$$y(k) = \boldsymbol{\varphi}_0^T(k) \boldsymbol{\theta} + e(k) \quad (33)$$

In relation (33) the vector  $\boldsymbol{\varphi}_0(k)$  depends on the immeasurable quantity  $e(i)$  ( $i = k-1, \dots, k-r$ ). The standard procedure in identification is to replace  $e(k)$  with the estimated prediction error. From that fact we have the following information vector

$$\boldsymbol{\varphi}^T(k) = [\mathbf{h}(k), \varepsilon_p(k-1), \dots, \varepsilon_p(k-r)] \quad (34)$$

Using the well known methodology from the system identification [23] and relations (20), (21) and (34) one can obtain the following recursive algorithm

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \alpha(k) \mathbf{P}(k) \boldsymbol{\varphi}(k) \varepsilon_p(k) \quad (35)$$

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1) \boldsymbol{\varphi}(k) \boldsymbol{\varphi}^T(k) \mathbf{P}(k-1)}{\alpha^{-1}(k) + \boldsymbol{\varphi}^T(k) \mathbf{P}(k-1) \boldsymbol{\varphi}(k)} \quad (36)$$

$$\alpha^{-1}(k) = \begin{cases} \sigma_N^2, & |\varepsilon_p(k)| \leq k_\varepsilon \\ \frac{|y(k) - \hat{\boldsymbol{\theta}}^T(k-1) \boldsymbol{\varphi}(k)|}{k_\varepsilon}, & |\varepsilon_p(k)| > k_\varepsilon \end{cases} \quad (37)$$

The algorithm (35) – (37) is new. In relation (37)  $k_\varepsilon$  is a Huber's constant. Three next remarks are important for simulations.



**Remark 2.** For the Gaussian disturbance  $\psi(x) = x$ . In that case the recursive algorithm has the form

$$\hat{\theta}(k) = \hat{\theta}(k-) + \frac{1}{\sigma_N^2} \mathbf{P}(k) \boldsymbol{\varphi}(k) \varepsilon_p(k) \quad (38)$$

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1) \boldsymbol{\varphi}(k) \boldsymbol{\varphi}^T(k) \mathbf{P}(k-1)}{\sigma_N^2 + \boldsymbol{\varphi}^T(k) \mathbf{P}(k-1) \boldsymbol{\varphi}(k)} \quad (39)$$

**Remark 3.** Recently Filipovic (2019) proposed the next robust recursive algorithm which has the form

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \mathbf{P}(k) \boldsymbol{\varphi}(k) \psi(\varepsilon_p(k)) \quad (40)$$

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1) \boldsymbol{\varphi}(k) \boldsymbol{\varphi}^T(k) \mathbf{P}(k-1)}{\alpha^{-1}(k) + \boldsymbol{\varphi}^T(k) \mathbf{P}(k-1) \boldsymbol{\varphi}(k)} \quad (41)$$

$$\alpha^{-1}(k) = \begin{cases} \sigma_N^2, & |\varepsilon_p(k)| \leq k_\varepsilon \\ h, & |\varepsilon_p(k)| > k_\varepsilon \end{cases} \quad (42)$$

$$h = \frac{2\sigma_N^5}{k_\varepsilon^3 \sqrt{2\pi}} \exp\left\{-\frac{k_\varepsilon^2}{2\sigma_N^2}\right\} \quad (43)$$

$$\psi(x) = \max\{-k_\varepsilon, \min(k_\varepsilon(x), x)\} \quad (44)$$

In the simulation section a comparison of the algorithm proposed in this paper and the algorithm (38)-(39) and algorithm (40) – (44) will be made.

**Remark 4.** Simulations show that the robust recursive algorithm is more sensitive, in comparison to the linear algorithm, to initial conditions. Because, the general approach to set  $\hat{\theta}(0) = 0$ ,  $\mathbf{P}(0) = \rho \mathbf{I}$ ,  $\rho = 10^4$  is necessary to replace on next mode

a) NFIR and NARX models

In this case, linear recursive algorithms have initial conditions [23]

$$\mathbf{P}(k_0) = \left[ \sum_{k=1}^{k_0} \boldsymbol{\varphi}(k) \boldsymbol{\varphi}^T(k) \right]^{-1}$$

$$\hat{\theta}(k_0) = \mathbf{P}(k_0) \sum_{k=1}^{k_0} \boldsymbol{\varphi}(k) y(k)$$

where  $k_0 = \dim \boldsymbol{\varphi}(k) = \dim \hat{\theta}(k)$ . Now  $\mathbf{P}(k_0)$  and  $\hat{\theta}(k_0)$  are initial conditions for the robust recursive algorithm

## b) NARMAX model

In this case it is impossible to apply the above methodology owing to the presence of a prediction error in the information vector  $\varphi(\cdot)$ . The simplest way, which in practice works well, is as follows

- i) For the initial condition  $\hat{\theta}(0) = 0$  and  $P(0) = \rho I$ ,  $\rho = 10^4$  the linear algorithm works  $k_0 = 20 \div 30$  iterations
- ii) Given values for  $\hat{\theta}(k_0)$  and  $P(k_0)$  now are the initial condition for the robust algorithm.

It is possible to use the described methodology, also, for NFIR and NARX models

## 5. SIMULATION STUDY

The properties of the proposed robust recursive algorithms for identification of Hammerstein systems is considered on the simulation level. We will consider the next Hammerstein model

$$v(k) = f(u(k)) = u(k) + d_2 u^2(k) + d_3 u^3(k) = u(k) + 0.65u^2(k) + 0.35u^3(k)$$

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} = 1 - 0.58q^{-1} + 0.6q^{-2}$$

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} = 0.8q^{-1} - 0.5q^{-2}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} = 1 - 0.4q^{-1}$$

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  $N(m, \sigma^2)$  is the Gaussian distribution with mean  $m$  and variance  $\sigma^2$ . In all simulations it is supposed that  $\sigma_1^2 = 1$  and  $\sigma_2^2 = 100$ .

The form of estimation error is  $E_k = \ln \left\| \hat{\theta}(k) - \theta \right\|^2$

We will consider three types of errors

ELS – for linear algorithm (38) – (39)

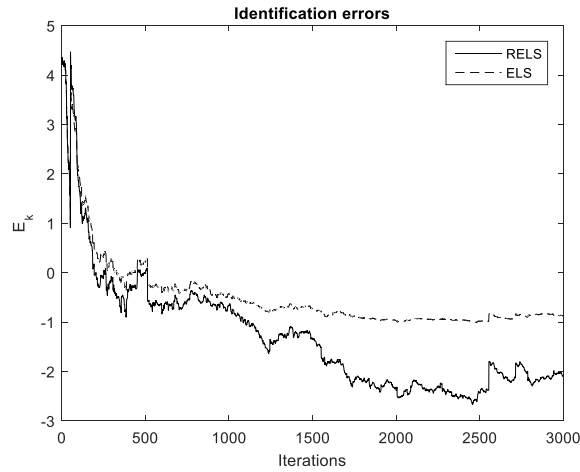
RELS – for algorithm (35) – (37)

RELS1 – for algorithm (40) – (44)

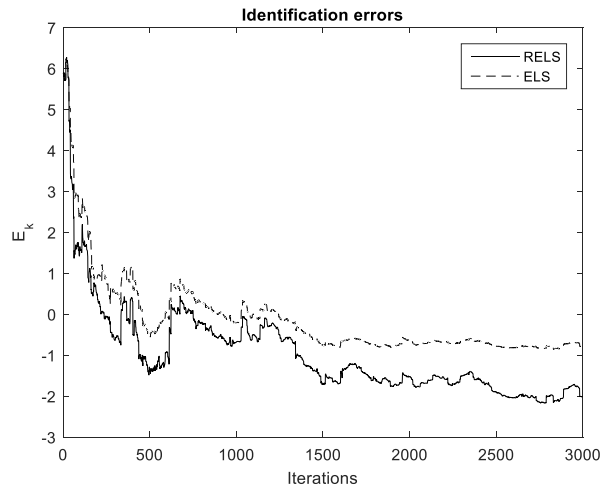
The simulations can be divided in two parts.

A) Comparison of algorithms (38) – (39) and (35) – (37)

The following degrees of contamination are considered  $\varepsilon = 0.01; 0.05$ ; The Huber's parameter is  $k_\varepsilon = 2$ .



**Fig. 1** Comparison of RELS and ELS for  $\varepsilon = 0.01$



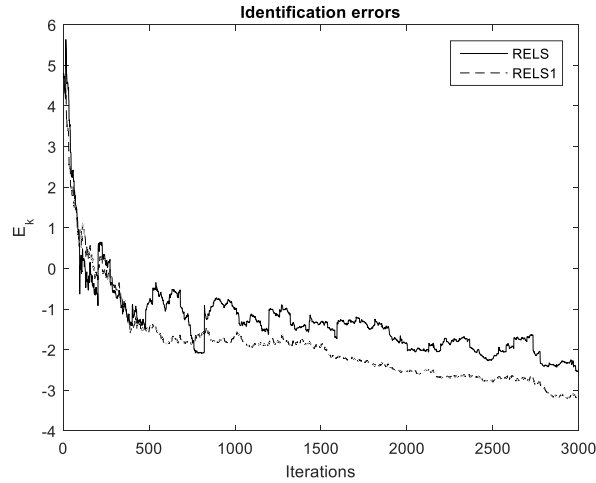
**Fig. 2** Comparison of RELS and ELS for  $\varepsilon = 0.05$

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

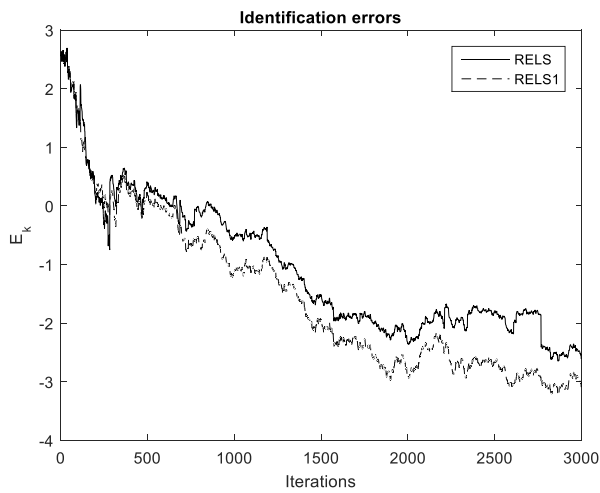
B) Comparison of algorithms (35) – (37) and (40) – (44)  
the next degrees of contamination are considered

$$\varepsilon = 0.01; \quad \varepsilon = 0.05$$

The Huber's parameter is  $k_\varepsilon = 2$ .



**Fig. 3** Comparison of RELS and RELS1 for  $\varepsilon = 0.01$



**Fig. 4** Comparison of RELS and RELS1 for  $\varepsilon = 0.05$

From the last two figures it follows that the robust algorithm (41) – (47) is better than the algorithm proposed in the paper (36) – (38).

The global conclusion, based on simulations, is that the algorithm, proposed in the paper, is superior in comparison with linear algorithms but is somewhat weaker in comparison with one class of the recently proposed robust algorithm. The last fact is the price for the Weiszfeld transformation.

## 6. CONCLUSION

The paper considers the design of a robust recursive algorithm based on applications of robust statistics and the Weiszfeld algorithm. The least favourable probability density on a class of distributions gives the loss function which consists of the  $l_2$  - norm part (Gaussian distribution) and the  $l_1$  - norm part (Laplace distribution). By using the Weiszfeld algorithm the above mentioned two criteria are degenerate in the generalized  $l_2$  - norm criterion. Robustness is included in the weighting factor of the  $l_2$  - norm criterion. Simulations show that the algorithm, proposed in the paper, is superior in comparison with linear algorithms and is somewhat weaker in comparison with the robust algorithm based on the  $l_1 - l_2$  - criterion and that is the price for the Weiszfeld transformation. Further research is related to the recursive algorithm design for multivariable systems and the case when identification criterion is the  $l_1$  - norm functional.

## REFERENCES

- [1] T. Soderstrom, P. Stoica, System Identification, Prentice-Hall, New Jersey, 1989.
- [2] R. Pearson, Exploring Data in Engineering, the Science and Medicine, Oxford University Press, Oxford, 2011.
- [3] P. Huber, E. Ronchetti, Robust Statistics, Willey, New York, 2009.
- [4] H. Kodomama, B. Huang, R. Ranjan, Y. Zhao, R Tan, N. Sammakneyad, "Approaches to robust process identification: A review and tutorial of probability methods," Journal of Process Control, vol. 66, no.6 , pp. 68-83, 2018. [Online]. Available: <https://doi.org/10.1016/j.jprocont.2018.02.011>
- [5] V. Filipovic, "Outlier robust stochastic approximation algorithm for identification of MIMO Hammerstein models," Nonlinear Dynamics, vol. 90, no. 2, pp. 1427-1441, 2017. [Online]. Available: <http://DOI10.1007/s11071-017-3736-2>
- [6] V. Filipovic, "A global convergent outlier robust adaptive predictor for MIMO Hammerstein models," International Journal of Robust and Nonlinear Control, vol. 27, no.16, pp. 3350-3371, 2017. [Online]. Available: <https://doi.org/10.1002/rnc.3705>
- [7] V. Filipovic, "Recursive identification of block-oriented nonlinear systems in the presence of outliers," Journal of Process Control, vol. 78, no.6, pp. 1-12, 2019. [Online]. Available: <https://doi.org/10.1016/j.jprocont.2019.03.015>
- [8] M. Sugiyama, Introduction to Statistical Machine Learning, Morgan Kaufman, San Francisco, 2016.
- [9] M. Ghanmi, M. Elloumi, H. Salhi, S. Kamoun, "A recursive hierarchical parametric estimation algorithm for nonlinear systems described by Wiener – Hammerstein models," Asian J. Control, vol. 21, no. 1, pp. 1-10, 2019. [Online]. Available: <https://doi.org/10.1002/asjc.2029>
- [10] F. Giri, E.W. Bai, Block-Oriented Nonlinear System Identification, Springer, Berlin, 2010.
- [11] E. Weiszfeld, "Sur le point pour lequel la somme des distances de n points donnees est minimum," Tohoku Math. J., vol. 43, First series, pp.355-386, 1937. [Online]. Available: <http://pdf.jst.go.jp>
- [12] E. Weiszfeld, "On the point for which the sum of distances to n given points in minimum," Anu. Oper. Res., vol. 167, no.1, pp. 7-41, 2009. Translate from the French original (Tohoku Math. J. vol. 43, First series, pp. 355-386, 1937. [Online]. Available: <http://DOI10.1007/s10479-008-0352-z>
- [13] P.A. Akimov, A.J. Matasov, "Level of non-optimality of the Weiszfeld algorithm in the least-modules method," Automation and Remote Control, vol. 71, no. 2, pp. 172-184, 2010. [Online]. Available: <http://doi:10.1134/S0005117910020025>
- [14] H.A. Eiselt, H Marianov, Foundations of Location Analysis, Springer, Berlin, 2011.

- [15] L. Mallozzi, E. D'Amato, P.M Pardalos, (Eds.), *Spatial Interaction Models. Facility Location using Game Theory*, Springer, Berlin, 2017.
- [16] Beck, S. Sabach, "Weiszfeld method: old and new results," *Journal of Optimization Theory and Applications*, vol. 164, no. 1, pp. 1-40, 2015. [Online]. Available: <http://DOI10.1007/s10957-011-0586-7>
- [17] Weber, *Theory of the Location of Industries*. University of Chicago Press, Chicago, 1909. P. Bloomfield, W.L. Steiger, *Least Absolute Deviations: Theory, Applications and Algorithms*, Birkhauser, Basel, 1983.
- [18] V.I. Mudrov, V.L. Kushko, *Methods of Measurements Processing*, (in Russian), Radio Svjaz, Moscow, 1983.
- [19] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
- [20] H. Kuhn, A note of Fermat problem. *Math. Programming*, vol. 4, no.1, pp. 98-107, 1973. [Online]. Available: <http://DOI:10.1007/bfo1584648>
- [21] V. Hutson, J. Pym, M. Cloud, *Application of Functional Analysis and Operator Theory*, Elsevier, London, 2005.
- [22] L. Ljung, T. Soderstrom, *Theory and Practice of Recursive Identification*, MIT Press, Massachusetts, 1983.
- [23] J.T. Barron, A general and adaptive robust loss function, arXiv: 1701.03077v10.