Adaptive Input Design for Robust Identification of Output-constrained OE Models

Vladimir Stojanovic[1*], Novak Nedic[1], Dragan Prsic[1]

[1]Faculty of Mechanical and Civil Engineering in Kraljevo, Department of Automatic Control, Robotics and Fluid Technique, University of Kragujevac (Serbia)

A robust identification of output error (OE) models with optimal input design for a case of constrained output variance is considered in this paper. In a case when observations have Gaussian mixture distributions, it is shown that the proposed robust algorithm for identification of OE models with constrained output, which is based on Huber's function, will give more accurate results in relation to the classical linear algorithm. In a form of the theorem, it is shown that an optimal input signal can be achieved by a minimum variance controller whose reference is a white noise. The essential problem is that the optimal input depends on the system parameters to be identified. In order to overcome this problem, a two-stage adaptive procedure is proposed, where iterations are alternately carried out between parameter estimation and experiment design using the current parameter estimates. It is shown that such obtained excitation signals result in a significant increasing in a convergence rate. Theoretical results are illustrated by simulations.

Keywords: input design, output error model, constrained output, robust identification algorithm

1. INTRODUCTION

The area, which deals with obtaining the mathematical model of the process, remains vibrant, as shown by recent research [1,2]. The main task of the theory of identification is the extraction of maximum information from the measurements that are available. This requirement is realized by optimal experiment design [3]. The basic approach consists in minimizing the scalar function of the Fisher information matrix [4].

The key problem in the optimal input design is that the optimal input depends on the unknown system properties, which need to be identified. Namely, the Fisher information matrix typically depends on system parameters. There are two basic approaches to overcome this problem. The first approach is based on robust optimal experiment design. In this case, the procedure is slightly sensitive to the uncertainty of a priori information about the system [5]. The basic a priori knowledge of the system can be obtained using a nonparametric frequency method [6]. The new results in this area cover the case of a finite number of model parameters and a very large number of measurements. The second approach is based on adaptation. One such, two-stage procedure has been proposed in [7]. In the first stage, in a short time interval, the data are collected using PRBS input. Based on these data, a system model is identified, and that is an initial model for optimal input design. In the second stage, the obtained input signal, by using a minimum variance controller and a stochastic reference, is used to generate a new data set. Adaptive input design for the ARX models has been discussed in [4].

In many practical cases, constraints on the fluctuation of input and/or output signals are very important [8]. For example, in the industrial production, product quality must be within certain limits (constraints on the fluctuation of the output signal).

If the constraint is related to the variance of the output signal, it is shown that the experiment design is D–optimal and that the input signal is generated using a minimum variance controller together with an external stochastic signal [8].

This paper considers the optimal experiment design for output error (OE) models. There is a constraint on the output power. When the corresponding noise is modelled as the Gaussian stochastic process, it is demonstrated in [9] that the presence of feedback is very important. Here, those results, using the Fisher information matrix, are extended on the case when the measurement noise is non-Gaussian. Justification of this approach was confirmed in practice [10]. Namely, in measurements there are rare, inconsistent observations with the largest part of population of observations (outliers). The presence of outliers can considerably degrade the performance of linearly recursive algorithms based on the assumptions that measurements have a Gaussian distribution. Therefore, synthesis of robust algorithms is of primary interest. The synthesis is based on Huber's theory of robust statistics [10]. Simulation results have shown that the proposed robust output error (OE) parameter estimation algorithm, based on the minimization of a robust criterion, will give more accurate results in relation to the conventional linear estimation algorithm, based on the recursive least squares (RLS) method.

It is shown that the optimal input signal can be obtained by a minimum variance controller whose reference is a white noise sequence with known variance. In order to be able to implement the algorithm, adaptive approach is applied. A direct adaptive minimum variance controller is used. The algorithm has two stages. In the first stage, the process parameters are estimated. In the second stage, based on thus obtained parameters, it is formed the minimum variance controller that generates the input signal of the process by which the identification is made. Because the reference signal is in the form of white noise, parameter estimation is consistent. The paper's results are supported by simulations.

*Corresponding author: Dositejeva 19, Kraljevo (Serbia), stojanovic.v@mfkv.kg.ac.rs

2. ROBUST IDENTIFICATION ALGORITHM FOR OE MODELS

The general form of the OE model is

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

where $u(k)$, $y(k)$ and $e(k)$ are input, output and stochastic noise, respectively. Polynomials $A(q^{-1})$ and $B(q^{-1})$ have the form:

$$\begin{aligned}A(q^{-1})&=1+a_1q^{-1}+\ldots+a_nq^{-n}\\ B(q^{-1})&=b_1q^{-1}+\ldots+b_mq^{-m}\end{aligned} \tag{2}$$

where q^{-i}, $i\in N$ denotes the backward shift operator i.e. $q^{-i}x(k)=x(k-i)$, while $b_i(i=1,\ldots,m)$ and $a_i(i=1,\ldots,n)$ are unknown parameters.

Practical and theoretical studies have shown that in a stochastic model of the system there are some observations that are inconsistent with the largest part of the population (outliers) [10], and that is why the disturbance (measurement noise) $e(k)$ in the model (1) is a non-Gaussian. Hence, the probability density function of the disturbance belongs to approximately normal distribution class:

$$\mathcal{P}_\varepsilon=\left\{p(e):p(e)=(1-\varepsilon)p_1(e)+\varepsilon p_2(e)\right\} \tag{3}$$

in which

$$p_1(e)\sim\mathcal{N}(0,\sigma_1^2),\ p_2(e)\sim\mathcal{N}(0,\sigma_2^2),\ \sigma_2^2\gg\sigma_1^2 .$$

In other words, the probability density function $p(e)$ represents a mixture of normal (Gaussian) distributions where σ_1^2 and σ_2^2 denote variances. The parameter $0\le\varepsilon<1$ is called the degree of contamination.

Let us introduce an auxiliary model

$$y_M(k)=\frac{B(q^{-1})}{A(q^{-1})}u(k), \tag{4}$$

or in the following form:

$$\begin{aligned}y_M(k)=&-a_1y_M(k-1)-\ldots-a_ny_M(k-n)+\\&+b_1u(k-1)+\ldots+b_mu(k-m)\end{aligned} \tag{5}$$

Since true values of parameters $a_i(i=1,\ldots,n)$ and $b_i(i=1,\ldots,m)$ are unknown, the output of the auxiliary model is calculated using estimates of parameters:

$$\begin{aligned}\hat{y}_M(k)=&-\hat{a}_1\hat{y}_M(k-1)-\ldots-\hat{a}_n\hat{y}_M(k-n)+\\&+\hat{b}_1u(k-1)+\ldots+\hat{b}_mu(k-m)\end{aligned} \tag{6}$$

Let $\hat{\theta}(k)$ be the estimated vector of parameters, and $\varphi(k)$ be the observation vector at the moment k :

$$\begin{aligned}&\hat{\theta}=[\hat{a}_1,\ldots,\hat{a}_n,\hat{b}_1,\ldots,\hat{b}_m]^T,\\&\varphi(k)=[-\hat{y}_M(k-1)\ldots-\hat{y}_M(k-n),u(k-1)\ldots u(k-m)]^T\end{aligned} \tag{7}$$

At the moment k, before the estimate $\hat{\theta}(k)$ is known, the prediction of the model is:

$$\hat{y}_M(k)=\hat{\theta}^T(k-1)\varphi(k) . \tag{8}$$

The problem of identification of the system described by (1) can be considered as the task of finding the vector $\hat{\theta}$, such that the mean square criterion:

$$\mathcal{J}(\hat{\theta})=E\left\{\varepsilon^2(k)\right\},\varepsilon(k)=y(k)-\hat{y}_M(k) \tag{9}$$

is minimized, where $E\{\cdot\}$ represents the mathematical expectation operator and $\varepsilon(k)$ is the prediction error.

Since the identification criterion (9) gives same weights to all residuals, these sporadic large observations (outliers) will have a significant influence on resulting parameter estimates. To achieve robustness we will consider the robust M-estimation criterion [10]:

$$\mathcal{J}_{\mathcal{R}}(\theta)=E\left\{\Phi\left[\varepsilon(k,\theta)\right]\right\} \tag{10}$$

in which Φ represents a robust loss function:

$$\Phi(\cdot)=-\log p^*(\cdot) . \tag{11}$$

By using Huber's min-max approach, it is possible to find the least favorable probability density function $p^*(\cdot)$ on a class of approximately normal distributions (3):

$$p^*\left(e(\mathrm{k})\right)=\begin{cases}\dfrac{1-\varepsilon}{2\pi\sigma_1}\exp\left\{-\dfrac{e^2(\mathrm{k})}{2\sigma_1^2}\right\} & |e(\mathrm{k})|\le k_\varepsilon\\ \dfrac{1-\varepsilon}{2\pi\sigma_1}\exp\left\{-\dfrac{k_\varepsilon}{\sigma_1^2}\left(|e(\mathrm{k})|-\dfrac{k_\varepsilon}{2}\right)\right\} & |e(\mathrm{k})|>k_\varepsilon\end{cases} \tag{12}$$

where k_ε is the Huber function parameter.

The empirical robust criterion on the observed interval, obtained from the relation for sufficiently large *k*, has the form:

$$\mathcal{J}_k(\theta)=\frac{1}{k}\sum_{t=1}^{k}\left\{\Phi(\varepsilon(i))\right\} . \tag{13}$$

Expanding $\mathcal{J}_k(\theta)$ in the vicinity of the preceding estimate $\hat{\theta}(\mathrm{k}-1)$ in Taylor series, one obtains:

$$\begin{aligned}\mathcal{J}_k(\theta)=&\mathcal{J}_k\left(\hat{\theta}(\mathrm{k}-1)\right)+\nabla_\theta\mathcal{J}_k\left(\hat{\theta}(\mathrm{k}-1)\right)\left[\theta-\hat{\theta}(\mathrm{k}-1)\right]+\\&+O\left(\left\|\theta-\hat{\theta}(\mathrm{k}-1)\right\|^2\right)\end{aligned} \tag{14}$$

where

$$\lim_{\|x\|\to\infty}\frac{O\left(\|x\|\right)}{\|x\|}=0 . \tag{15}$$

By minimizing the expression (14), it is obtained:

$$\begin{aligned}\hat{\theta}(\mathrm{k})=&\hat{\theta}(\mathrm{k}-1)-\left[k\nabla_\theta^2\mathcal{J}_k\left(\hat{\theta}(\mathrm{k}-1)\right)\right]^{-1}\left[k\nabla_\theta\mathcal{J}_k\left(\hat{\theta}(\mathrm{k}-1)\right)\right]+\\&+O\left(\left\|\theta-\hat{\theta}(\mathrm{k}-1)\right\|\right)\end{aligned} \tag{16}$$

A recursive form of the robust criterion (13) can be obtained as $\theta = \hat{\theta}(\mathrm{k}-1)$

$$k\mathcal{J}_k(\theta) = (k-1)\mathcal{J}_{k-1}(\theta) + \Phi(\varepsilon(k)). \tag{17}$$

By differentiating the last relation twice one can obtain:

$$k\nabla_\theta^2 \mathcal{J}_k(\theta) = (k-1)\nabla_\theta^2 \mathcal{J}_{k-1}(\theta) + \Psi'(\varepsilon(k))\varphi(k)\varphi^T(k) \tag{18}$$

where $\Psi(\cdot) = \Phi'(\cdot)$, $\Psi(\cdot): R^1 \to R^1$.

Furthermore, the following assumptions will be used:

a) The estimate $\hat{\theta}(\mathrm{k})$ is in the vicinity of the estimate $\hat{\theta}(\mathrm{k}-1)$

b) The estimate $\hat{\theta}(\mathrm{k}-1)$ is optimal at the instant k-1.

After replacing θ with $\hat{\theta}(\mathrm{k}-1)$ in the relation (18), one can obtain:

$$\begin{aligned} k\nabla_\theta^2 \mathcal{J}_k(\hat{\theta}(\mathrm{k}-1)) &= (k-1)\nabla_\theta^2 \mathcal{J}_{k-1}(\hat{\theta}(\mathrm{k}-1)) + \\ &+\Psi'(\varepsilon(k))\varphi(k)\varphi^T(k) \end{aligned} \tag{19}$$

From the assumption a) it follows

$$\nabla_\theta^2 \mathcal{J}_k(\hat{\theta}(\mathrm{k})) \cong \nabla_\theta^2 \mathcal{J}_k(\hat{\theta}(\mathrm{k}-1)) \tag{20}$$

Based on this, the relation (19) takes the form

$$\begin{aligned} k\nabla_\theta^2 \mathcal{J}_k(\hat{\theta}(\mathrm{k}-1)) &= (k-1)\nabla_\theta^2 \mathcal{J}_{k-1}(\hat{\theta}(\mathrm{k}-2)) + \\ &+\Psi'(\varepsilon(k))\varphi(k)\varphi^T(k) \end{aligned} \tag{21}$$

Based on the assumption a) it also follows

$$O\left(\left\|\hat{\theta}(\mathrm{k}) - \hat{\theta}(\mathrm{k}-1)\right\|\right) = 0 \tag{22}$$

By introducing the notation $\overline{R}(\mathrm{k}) = k\nabla_\theta^2 \mathcal{J}_k(\hat{\theta}(\mathrm{k}-1))$ from relations (16) and (21) one can obtain:

$$\hat{\theta}(\mathrm{k}) = \hat{\theta}(\mathrm{k}-1) - \overline{R}^{-1}(\mathrm{k})\left[k\nabla_\theta \mathcal{J}_k\left(\hat{\theta}(\mathrm{k}-1)\right)\right] \tag{23}$$

$$\overline{R}(\mathrm{k}) = \overline{R}(\mathrm{k}-1) + \Psi'(\varepsilon(\mathrm{k}))\varphi(k)\varphi^T(k) \tag{24}$$

From the assumption b) it follows $\nabla_\theta \mathcal{J}_{k-1}(\hat{\theta}(\mathrm{k}-1)) = 0$. Based on this condition, and if $\theta = \hat{\theta}(\mathrm{k}-1)$ is put in the relation (25), one obtains:

$$k\nabla_\theta \mathcal{J}_k(\hat{\theta}(\mathrm{k}-1)) = -\Psi(\varepsilon(k))\varphi(k) \tag{25}$$

Finally, based on relations (23) - (25), using the notation $P(\mathrm{k}) = \overline{R}^{-1}(\mathrm{k})$ and applying the matrix inversion lemma, one can obtain the definitive form of a recursive algorithm:

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

$$\mathrm{P(k)} = P(\mathrm{k}-1) - \frac{P(\mathrm{k}-1)\varphi(k)\varphi^T(k)P(\mathrm{k}-1)}{\left[\Psi'(\varepsilon(\mathrm{k}))\right]^{-1} + \varphi^T(k)P(\mathrm{k}-1)\varphi(k)} \tag{27}$$

$$\varepsilon(\mathrm{k}) = y(\mathrm{k}) - \hat{\theta}^T(\mathrm{k}-1)\varphi(k) \tag{28}$$

$$\Psi(x) = \min\left\{|x|, k_\varepsilon\right\}\mathrm{sgn}(x) \tag{29}$$

$$\Psi'(x) = \begin{cases} 1 & |x| < k_\varepsilon \\ 0 & otherwise \end{cases} \tag{30}$$

The function defined by the relation (29) is the Huber function [10]. It is derived for a class of distributions (3). It is shown on Figure 1.

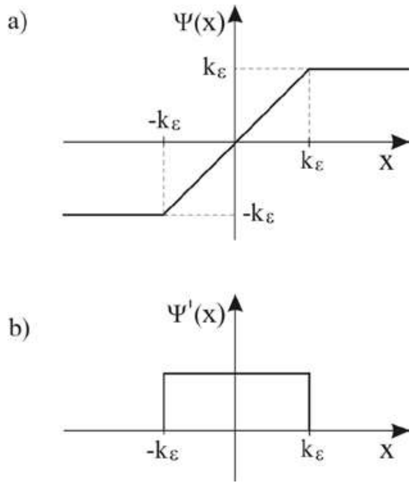

Figure 1: Nonlinear function of residuals
a) Huber's function
b) Derivative of Huber's function

3. OPTIMAL INPUT DESIGN FOR OE MODELS

Further, we will consider a special case of the model (1):

$$y(k) = \frac{b_1 q^{-1}}{A(q^{-1})}u(k) + e(k) = \frac{b_1}{A(q^{-1})}u(k-1) + e(k). \tag{31}$$

In this case the parameters b_1 and $a_i (i = 1,\ldots,n)$ are estimated.

The Fisher information matrix can be defined as [4]:

$$M = E_{Y|\beta}\left\{\left(\frac{\partial \log p(Y|\beta)}{\partial \beta}\right)^T \left(\frac{\partial \log p(Y|\beta)}{\partial \beta}\right)\right\} \tag{32}$$

Based on N measurements, the vector of outputs is $Y = \left[y(1) \ \ldots \ y(N)\right]^T$. Parameter vector β has the form:

$$\beta = \left[\theta^T \quad \sigma_e^2\right]^T \tag{33}$$

where σ_e^2 represents the variance of the noise $e(k)$, and $\theta^T = \left[a^T \quad b_1\right]$ in which $a^T = [a_1 \ \ldots \ a_n]$.

Since N is large, it is more convenient to work with the average value of the Fisher information matrix:

$$\overline{M} = \lim_{N \to \infty} \frac{1}{N} M \tag{34}$$

We shall principally use the determinant of the average information matrix as the design criterion leading to the following form of D-optimal criterion:

$$\mathcal{J} = -\log\det \overline{M}\,. \tag{35}$$

Minimization of the scalar cost function (35) relies upon the calculation of the matrix $\overline{M}$.

Lemma 1. [[11], pp. 232] Let $(\Omega,\mathcal{F},\mathrm{P})$ be a probability space. Suppose also that there is a subalgebra $\mathcal{F}_1$ $(\mathcal{F}_1 \subseteq \mathcal{F})$. Then

$$E\left\{E_{\xi|\mathcal{F}_1}\{\xi\}\right\} = E\{\xi\} \blacksquare \tag{36}$$

Lemma 2. [[8], pp. 244] Let $\overline{M}$ be a positive definite matrix with unit diagonal elements $\overline{M}_{ii} = 1$, $i = 1,\ldots,n$, then $\det\overline{M}$ achieves its maximum value for $\overline{M} = I$ $\blacksquare$

We will now formulate the optimal input in the form of the theorem.

Theorem 1. Suppose that for OE model (31), the following conditions are fulfilled:

1° Stochastic noise $e(k)$ represents a zero mean Gaussian mixture with variance σ_e^2,

2° A constraint on the output is $E\left\{y^2(k)\right\} \le W$, $W \in (0,\infty)$.

Then the criterion $-\log\det\overline{M}$ achieves its minimum value if the system input $u(k)$ is generated by the minimum variance controller which reference is a zero mean white noise sequence $\{\eta(k)\}$ with probability density function

$$p(\eta) = \frac{b_1}{\sqrt{2\pi W}} e^{-\frac{(\eta b_1)^2}{2W}} \quad \blacksquare \tag{37}$$

Proof: Let us define the residual sequence as:

$$\varepsilon(k) = y(k) + a_1 y_M(k-1) + \ldots + a_n y_M(k-n) - b_1 u(k-1) \tag{38}$$

Since we consider N observations, the likelihood function would be:

$$p(Y|\beta) = \prod_{k=1}^{N}\left((1-\varepsilon)p_1\left(y(k)\right) + \varepsilon p_2\left(y(k)\right)\right). \tag{39}$$

After some calculations, according to **Lemma 1**, the mean value of the Fisher information matrix can be expressed as:

$$\begin{aligned}\overline{M} = {} & \frac{1-\varepsilon}{\sigma_1^2} E\left\{\left(\frac{\partial\varepsilon(k)}{\partial\beta}\right)^T\left(\frac{\partial\varepsilon(k)}{\partial\beta}\right) + \frac{1}{2\sigma_1^4}\left(\frac{\partial\sigma_1^2}{\partial\beta}\right)^T\left(\frac{\partial\sigma_1^2}{\partial\beta}\right)\right\} + \\ & + \frac{\varepsilon}{\sigma_2^2} E\left\{\left(\frac{\partial\varepsilon(k)}{\partial\beta}\right)^T\left(\frac{\partial\varepsilon(k)}{\partial\beta}\right) + \frac{1}{2\sigma_2^4}\left(\frac{\partial\sigma_2^2}{\partial\beta}\right)^T\left(\frac{\partial\sigma_2^2}{\partial\beta}\right)\right\}\end{aligned} \tag{40}$$

From relations (33), (38), and (40) one can obtain

$$\overline{M} = \frac{1}{\sigma_e^2} E\left[\begin{array}{ccc:c:c} y_M(k-1)^2 & \ldots & y_M(k-1)\cdot y_M(k-n) & -u(k-1)y_M(k-1) & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ y_M(k-n)\cdot y_M(k-1) & \ldots & y_M(k-n)^2 & -u(k-1)y_M(k-n) & 0 \\ \hdashline -u(k-1)y_M(k-1) & \ldots & -u(k-1)y_M(k-n) & u(k-1)^2 & 0 \\ \hdashline 0 & \ldots & 0 & 0 & 1/2\sigma_e^2 \end{array}\right] \tag{41}$$

where the sub-matrices correspond to the partitioning of the parameter vector β into $\left[a^T \;\vdots\; b_1 \;\vdots\; \sigma_e^2\right]^T$, in which σ_e^2 represents the variance of Gaussian mixture noise.

The relation (41) can be written in a form that is more compact:

$$\overline{M} = \frac{1}{\sigma_e^2}\left[\begin{array}{c:c:c} A & B & 0 \\ \hdashline B^T & C & 0 \\ \hdashline 0 & 0 & 1/2\sigma_e^2 \end{array}\right] \tag{42}$$

The following task is to determine elements of the matrix $\overline{M}$ in the relation (42).

Step 1 (Determining the matrix A)

Let us define

$$E\left\{y_M(k-i)y_M(k-j)\right\} \triangleq \rho_{|i-j|} \tag{43}$$

Based on the relation (43) the matrix A can be presented in the following form:

$$A = \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & \rho_0 \end{bmatrix} \tag{44}$$

From relations (4) and (31) it follows:

$$y(k) = -a_1 y_M(k-1) - \ldots - a_n y_M(k-n) + b_1 u(k-1) + e(k) \tag{45}$$

After multiplying the relation (45) with $y_M(k-i)$, $i = 1,\ldots,n$ and applying the mathematical expectation operator to the obtained system of n equations, one can obtain:

$$V = -Af + Bb_1 \tag{46}$$

where $a^T = [a_1 \;\ldots\; a_n]$, $V = [\rho_1 \;\; \rho_2 \;\ldots\; \rho_n]^T$.

Finally, it follows from the relation (46) that:

$$B = \frac{1}{b_1}\left(Af + V\right) \tag{47}$$

Step 3 (Determining the scalar C)

From the relation (4) one can get the input signal $u(\cdot)$:

$$u(k-1) = \frac{1}{b_1}\left[y_M(k) + a_1 y_M(k-1) + \ldots + a_n y_M(k-n)\right] \tag{48}$$

Scalar C is defined as:

$$\begin{aligned} C = E\left\{u(k-1)^2\right\} = \frac{1}{b_1^2} E\Big[& y_M(k)^2 + a_1 y_M(k-1) y_M(k) + \ldots + a_n y_M(k-n) y_M(k) + \\ & + a_1 y_M(k) y_M(k-1) + a_1^2 y_M(k-1)^2 + \ldots + a_1 a_n y_M(k-1) y_M(k-n) + \\ & \vdots \\ & + a_n y_M(k-n) y_M(k) + a_n a_1 y_M(k-1) y_M(k-n) + \ldots + a_n^2 y_M(k-n)^2 \Big] \end{aligned} \tag{49}$$

After arranging individual terms of (49), the finally expression for the scalar C is given by:

$$C = E\left\{u(k-1)^2\right\} = \frac{1}{b_1^2}\left(\rho_0 + 2aV^T + a^T A a\right) \tag{50}$$

Since, all elements of the matrix $\overline{M}$ (relation (42)) are now known, one can obtain:

$$\det \overline{M} = \frac{1}{2\sigma_e^2} \det A \cdot \det\left(C - B^T A^{-1} B\right) \tag{51}$$

Now, it can be shown that:

$$\begin{aligned} & -\log\det\overline{M} = \\ & \log 2\sigma_e^2 - \log\left(\det A\right) - \log\left(\rho_0 - V^T A^{-1} V\right) + \log b_1^2. \end{aligned} \tag{52}$$

In accordance with the condition 2° of the **Theorem 1**, we have $\rho_0 = W$, so the diagonal elements of the matrix A are fixed. Based on **Lemma 2** it follows that $-\log\left(\det A\right)$ has a minimum value when A is the diagonal matrix. This means that $\rho_i = 0,\ i > 0$, which further gives $V = 0$. Because of this, the third term of the equation (52) has a minimum value $-\log\left(W\right)$. We finally get

$$\min\left\{-\log\det\overline{M}\right\} = \log\frac{2\sigma_e^2 b_1^2}{W^{n+1}} \tag{53}$$

It is necessary to note that $\rho_i = 0,\ i > 0$ is achieved when $\left\{y(k)\right\}$ is an uncorrelated sequence. This condition is fulfilled if the input signal is chosen in the following form:

$$u(k) = \frac{1}{b_1}\left[a_1 y_M(k) + \ldots + a_n y_M(k-n+1)\right] + \eta(k) \tag{54}$$

where $\eta(k)$ is a reference signal that represents white noise with variance W/b_1^2. The relation (54) represents the minimum variance controller for the model (31). This theorem is proved. ■

This result also shows that the optimal input design requires knowledge of the true system parameters. In practical conditions, however, such a requirement is contradictory because the optimal input design is performed in order to speed up the identification process. In other words, it is impossible that unknown system parameters are known a priori. In a real application this fact must be handled. In order to overcome this problem a two-stage adaptive procedure is proposed:

A) By using PRBS signal as input, through N_{init} iterations, the initial model of the process is determined,

B) After that, through N_{opt} iterations, adaptation is applied for the controller defined in **Theorem 1**.

In the section devoted to simulations, the proposed two-stage algorithm will be compared with the open loop system identification algorithm with the PRBS input signal.

4. SIMULATION RESULTS

The proposed two-stage identification algorithm is tested on the following OE model:

$$y(k) = \frac{0.5q^{-1}}{1 - 0.7q^{-1} + 0.5q^{-2}} u(k) + e(k) \tag{55}$$

The system identification example, is based on measured 1000 input-output data points obtained during the simulations. The measurement noise $e(k)$ has non-Gaussian distribution defined by:

$$\mathcal{P}_\varepsilon = \left\{p(e) : p(e) = (1-0.1)\mathcal{N}(0, 0.1) + 0.1\mathcal{N}(0, 10)\right\} \tag{56}$$

Figure 2 shows system input and corresponding system output when the robust two-stage identification algorithm is used. First 200 iterations, the system works in an open loop, with PRBS as the input signal. Then it switches feedback in the range of iterations 200-1000.

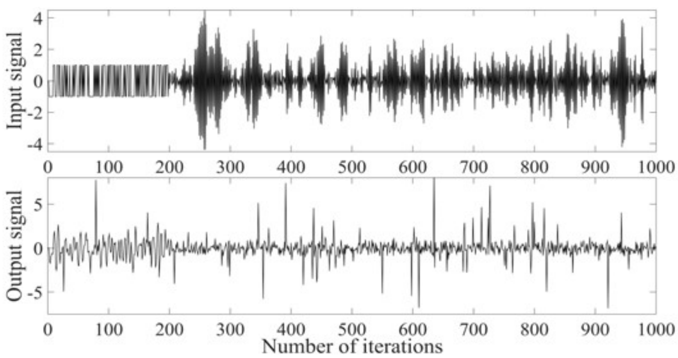

Figure 2: Adaptive input and corresponding output

Simulation results have illustrated significant increasing of accuracy in parameter estimates of OE model by using the robust identification algorithm in relation to the linear identification algorithm with PRBS [-1,1] as an input signal. Furthermore, it can be seen that the convergence rate of the robust algorithm is further increased by using the optimal input design, which further increases the practical value of proposed robust procedure.

Figures 3 and 4 show parameter estimates in the case where the output variance cannot be greater then $W = 0.5$.

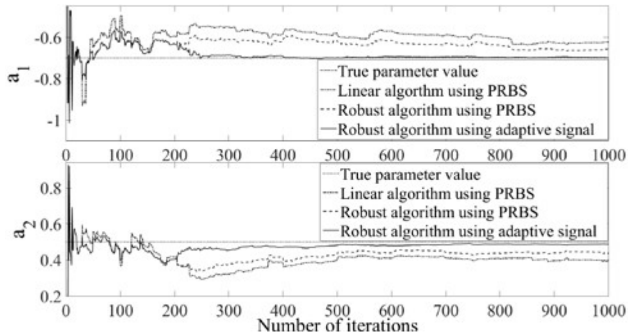

Figure 3: Estimates of parameters a_1 and a_2

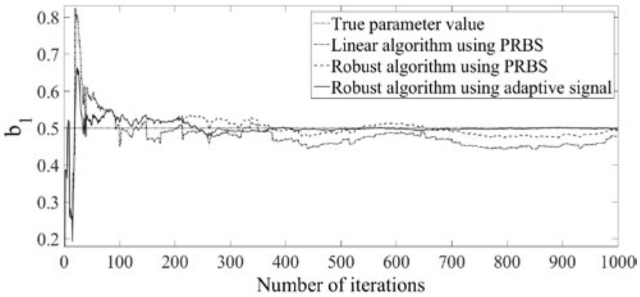

Figure 4: Estimate of parameter b_1

The simulation results are compared in terms of mean square errors (see Figure 5), defined by

$$MSE = \ln\left(\mathrm{E}\left\| \hat{\theta}(k) - \theta(k) \right\|^2 \right) \tag{57}$$

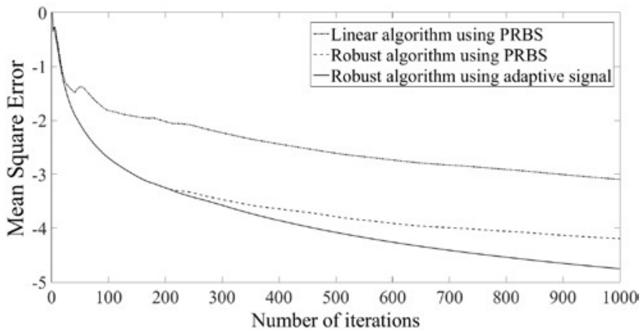

Figure 5: Mean square errors

Based on these figures, it can be concluded that experiment design increases the convergence speed of parameters to true values, keeping the given output variance W .

5. CONCLUSION

In this paper the optimal input design for robust identification of a special case of OE models, in the case of constrained output variance, has been considered.

Simulation results have shown that the proposed robust OE parameter estimation algorithm, produces more efficient parameter estimates in relation to the conventional linear algorithm.

Also, it is shown, that the optimal input is obtained by using the minimum variance controller and the stochastic reference signal. The adaptive two-stage procedure for generating the input signal is proposed. The initial model of the process is firstly obtained using PRBS input signal. In the second stage, the optimal input signal is generated by the minimum variance controller together with the stochastic reference. Simulation results have shown the superiority of the robust identification algorithm using proposed adaptive methodology for generating the optimal input signal.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the Serbian Ministry of Education, Science and Technological Development for supporting this paper through projects TR33026 and TR33027.

REFERENCES

[1] E.Zhang, R. Pintelon, "Identification of multivariable dynamic errors-in-variables system with arbitrary inputs", Automatica, Vol. 82, pp. 69-78, (2017)

[2] A. Janot, P.-O. Vandanjon, M. Gautier, "A revised Durbin-Wu-Hausman test for industrial robot identification", Control Engineering Practice, Vol.48, pp.52-62, (2016)

[3] M. Zarrop, "Optimal experiment design for dynamic system identification", Springer, Berlin, (1979).

[4] G.C. Goodwin, R.L. Payne, "Dynamic system identification: Experiment design and data analysis", Academic Press, NY, (1977).

[5] C.R. Rojas, J.C. Agüero, J.S. Welsh, G.C. Goodwin, A. Feruer, "Robustness in experiment design", IEEE Transactions on Automatic Control, Vol. 57(4), pp. 860-874, (2012).

[6] J.S. Welsh, G.C. Goodwin, "Finite sample properties of indirect nonparametric closed-loop identification", IEEE Transactions on Automatic Control, Vol. 47(8), pp.1277-1292, (2002).

[7] M. Barenthin, H. Jansson, H. Hjalmarsson, "Applications of mixed H_2 and $H\infty$ input design in identification", Proceedings of the 16th IFAC World Congress, Prague, July 3-8 2005, pp. 458-463, (2005).

[8] T.S. Ng, G.C. Goodwin, T. Söderström, "Optimal experiment design for linear systems with input-output constraints", Automatica, Vol. 13(6), pp.571- 577, (1977).

[9] V. Stojanovic, V. Filipovic, "Adaptive input design for identification of output error model with constrained output", Circuits, Systems, and Signal Processing, Vol.33(1), pp.97-113, (2014)

[10] P.J. Huber, "Robust Statistics", Wiley, NY, (1981).

[11] A.N. Shiryayev, "Probability", Nauka, Moscow, (1989).