X Triennial International SAUM Conference
on Systems, Automatic Control and Measurements
Niš, Serbia, November 10th-12th, 2010

Robust Kalman Filter as Parameter Estimator for Output Error Models

V. Filipovic and V. Stojanovic

Department of Energetics and Automatic Control, University of Kragujevac, Faculty of Mechanical Engineering Kraljevo, Dositejeva 19, 36000 Kraljevo, Serbia, E-Mail: vfilip@eunet.rs

***Abstract** - The data in industry are corrupted with stochastic noise. In the real situations data contain outliers which can create problems to linear algorithms. Because, some kind of prevention must be taken into account. So are developed robust procedures for parameters estimation. In this paper we shall consider output error model and for robust parameters estimation the Masreliez-Martin's robust filter is used. This filter is generalization of Kalman filter. In this paper we*

***(i)** eliminate the transformation factor*

***(ii)** nonlinear Masreliez-Martin prediction error transformation we replace with Huber function*

***(iii)** Fisher information is replaced with derivative of Huber's function*

***(iv)** generation of input signal (experiment design) is based on ideas from predictive control*

Also, the intensive simulations are performed.

***Key words:** Nongaussian noise, output error model, robust Kalman filter, experiment design*

I. INTRODUCTION

One of the problems that appears in the area of identification of industrial processes is the identification in the presence of stochastic disturbance. Practical studies show that disturbance, in general, has non-Gaussian distribution. It is an especially important case when appears inconsistent, in relation to the main part of population, observations (outliers). Probability distributions for this case are approximately normal (ε -contaminated) and they are the subject of intense research in mathematical statistics. For such a case, we consider the robust procedures for parameters estimation.

The reason for which we decided to use output error (OE) type recursive algorithm for parametric identification is that it can provide good performance when the plant output measurement is disturbed by noise. This can be explained by the fact that the output of this predictor does not directly depend upon the measured variables disturbed by noise, as the case with prediction error type predictor. The output of this predictor depends indirectly upon the measurements through the adaptation algorithm but this dependence can decrease in time by using a decreasing adaptation gain [1-2].

We will apply the robust filter theory for solving problem of the robust parameter estimation. In 1960, Kalman introduced an effective algorithm to realize the optimum filter for Gaussian processes, [3]. The Kalman filter works well, but it assumes that the system model and noise statistics are known. If any of these assumptions are violated then the filter estimates can degrade. This was noted early in the history of Kalman filtering [4-5].

Although the Kalman filter is the optimal linear filter, it is not the optimal filter in general for non-Gaussian noise. Noise in nature is often approximately Gaussian but with heavier tails, and the Kalman filter can be modified to accommodate these types of density functions [6-8].

Masreliez [6] found that the score function for the residual process plays an important role in obtaining the minimum variance estimator. However, the procedure to evaluate the score function involves convolution operations which are difficult to implement. Facing this difficulty, Masreliez and Martin [7] applied the influence function of min-max robust theory [9-10] to replace the score function. They first use the linear transformation to scale and symmetrize the density of the residual process and then operate on the result with an influence function to cut off the outliers in the noise distribution.

Nevertheless, several basic disadvantages are associated with the Masreliez-Martin filter:

- The estimator requires the unknown contaminating distribution to be symmetric. This is a stringent requirement, since in practice one would not expect departures from normality to be symmetric. Without the assumption of symmetry, the estimator may not be consistent.
- In general, there are relatively long periods of time in which noise is essentially Gaussian so that it is important to maintain full efficiency during these periods. Yet their estimator cannot work as well as the standard Kalman filter does in Gaussian noise.
- For various forms of measurement noise, the achievable optimum Cramer-Rao bound of point estimation (Kalman filtering without plant noise) is different. This estimator is optimal not in the efficiency-robust sense (approaches the Cramer-Rao bound), but in the min-max sense (minimizes the maximum variance) within those specific models such as the ε -contaminated family Φ_ε for a fixed mixing parameter ε [9] or a p-point family model Φ_p with a fixed p [10], respectively.

Because of these disadvantages, authors have decided to make some modification of the robust filter proposed by Masreliez and Martin [7].

Finally, the theory of experiment design is used in order to reduce the time needed for identification. The aim is to create the input signal for identification (excited signal) through a recursive relation for autocovariance function. Synthesis of autocovariance function is based on the ideas of predictive control, [11].

II. MODIFIED ROBUST KALMAN FILTER

We will consider a dynamic system whose $n \times 1$ dimensional state vector $x(k)$ $(k \geq 0)$ is estimated from scalar observations $y(1), \ldots, y(k)$, in which $x(k)$ and $y(k)$ satisfy the linear state and observation relations

$$x(k+1) = F(k)x(k) \tag{1}$$

$$y(k) = H(k)x(k) + \nu(k) \tag{2}$$

with $n \times n$ state transition matrix $F(k)$, $1 \times n$ observation matrix $H(k)$, and the scalar measurement noise $\nu(k)$. We can note that considered system has
As mentioned above, our attention is turned towards the case when observation noise has Non Gaussian distribution.

From [7], for situation of the linear model (1-2), the Masreliez-Martin's robust filter is defined by:

$$\hat{x}(k) = \overline{x}(k) + M(k)H^T(k)T^T(k)\psi_p[\nu(k)] \tag{3}$$

$$\overline{x}(k) = F(k-1)\hat{x}(k-1) \tag{4}$$

$$M(k) = F(k-1)P(k-1)F^T(k-1) \tag{5}$$

where the filter covariance, $P(k)$ satisfies

$$\begin{aligned} P(k) &= M(k) - \\ &- M(k)H^T(k)T^T(k)T(k)H(k)M(k)E_{f_0}\left\{\psi_p^{'}(\nu)\right\}, \end{aligned} \tag{6}$$

$T(k)$ is a linear transformation which insure the probability density function (pdf) of the transformed residual process

$$\nu(k) = T(k)\left[y(k) - H(k)\overline{x}(k)\right] \tag{7}$$

to be symmetric and all marginal pdf be members of the ε-contaminated family Φ_ε. Also is assumed that a priori state prediction error $x(k) - \overline{x}(k)$ is Gaussian for all k with zero mean and covariance $M(k)$. $E_{f_0}\{\cdot\}$ denotes expectation with respect to the least favorable pdf f_0, $\psi_p[\cdot]$ represents the vector influence function. Details may be found in [6-7].

Trying to avoid disadvantages specified in Introduction we made some modifications of this robust filter.

For symmetric residual process, the linear transformation T_k was asserted to exist. However, its existence strongly depends on the pdf of the residual process. Without *a priori* knowledge, it is difficult to implement. Even if this information is available, it is still not clear how to find T_k except for the scalar model. Adopting T_k =1, we eliminate the transformation factor.

The replacement of Fisher information $E_{f_0}\left\{\psi_p^{'}(\nu)\right\}$ by derivative of Huber's function enables great efficiency during periods of time in which noise is essentially Gaussian.

Based on extensive simulation results, it was observed that the robust filter modified in this way gives something better estimates in the event that nonlinear prediction transformation proposed by Masreliez and Martin [7] is replaced with Huber's function.

The proposed modifications of Masreliez-Martin's robust filter have been derived on the basis of approximations and somewhat heuristic reasoning.

Let's consider now the process model which is described by:

$$y(k,\theta) = \varphi^T(k)\theta + \nu(k) \tag{8}$$

with standard notations:

$$\begin{aligned} \varphi(k) &= [-y(k-1), \ldots, -y(k-n), u(k-1), \ldots, u(k-m)]^T \\ \theta &= [a_1, \ldots, a_n, b_1, \ldots, b_m]^T \end{aligned} \tag{9}$$

which denote the vector of input-output data and the vector of true constant parameter values, respectively. The density for observation noise $\nu(k)$ is represented by the Gaussian mixture:

$$\mathrm{p}(\nu(\mathrm{k})) = (1-\varepsilon)\mathrm{N}(\nu|0,\sigma_1^2) + \varepsilon\mathrm{N}(\nu|0,\sigma_2^2),\ 0 \leq \varepsilon < 1 \tag{10}$$

The output error adjustable predictor is described by:

$$\hat{y}(k) = \phi^T(k)\hat{\theta}(k) \tag{11}$$

where $\hat{y}(k)$ denotes the *a posteriori* output of the predictor, in which:

$$\begin{aligned} \phi(k) &= [-\hat{y}(k-1), \ldots, -\hat{y}(k-n), u(k-1), \ldots, u(k-m)]^T \\ \theta(k) &= [\hat{a}_1(k), \ldots, \hat{a}_n(k), \hat{b}_1(k), \ldots, \hat{b}_m(k)]^T \end{aligned} \tag{12}$$

The problem of robust output error recursive identification of a system described by (8) can be considered as the task of estimation the unknown the parameter vector θ in real time, on the basis of current input data and the *a posteriori* predicted output data. Since predicted output $\hat{y}(k)$ should converge asymptotically to the true output value $y(k)$, $\hat{y}(k)$ is an approximation of the output $y(k)$, which will improve as the time passes. However, if the linear regression model (11) can be cast into the form (1-2) by

$$\begin{aligned} \theta(k+1) &= \theta(k), \quad (=\theta) \\ y(k) &= \phi^T(k)\hat{\theta}(k) + \nu(k) \end{aligned} \tag{13}$$

then applying the robust Kalman filter to (13), with $F(t) = I$, $H(t) = \phi^T(t)$, robust output error parameter estimation of the system (8) can be done.

Generally, Kalman filter interpretation of linear parameter estimator can be seen in [12].

III. GENERATION OF INPUT SIGNAL

Optimal test signals are frequently specified in terms of their second order properties, e.g. autocovariance or spectrum. This leads to the problem of implementing a real signal with specified second order properties. In addition, it is usual that the input should also be constrained in its amplitude, therefore, the amplitude must lie in some interval.

Within the constraints of its amplitude, it is important to implement an input signal which has maximum power. It is of great importance in experiment design, where the quality of the estimation typically increases with the signal to noise ratio. If we choose an input with higher power, it is obviously that the signal to noise ratio is improved. Binary signals have precisely this desirable property: their power is maximum for given amplitude constraint.

As mentioned in introduction section, we utilize ideas from model predictive control to generate a binary waveform whose sampled autocovariance is as close as possible to some prescribed autocovariance, [11]. Heuristically speaking, the idea is to solve, for each time instant, a finite horizon optimisation problem to find the optimal set of the next, say, T values of the sequence such that the sampled autocovariance sequence so obtained is as close as possible (in a prescribed sense) to the desired autocovariance. One then takes the first term of this optimal set for the sequence, advances time by one step and repeats the procedure. The idea behind this procedure is thus closely related to finite alphabet receding horizon control, where receding horizon concepts are employed to control a linear plant whose input is restricted to belong to a finite set.

Before the algorithm begins, the user of the algorithm has to convert the desired autocovariance sequence $\left\{r_k^d\right\}_{k=0}^{\infty}$ into the non-central autocovariance of a {0, 1} sequence $\left\{\hat{r}_k^d\right\}_{k=0}^{\infty}$. Also, the user must choose three variables: N-the length of the signal to be generated, n-the number of lags $\left\{r_k^d\right\}_{k=0}^{\infty}$ to be compared to the corresponding lags of the sampled autocovariance sequence of the designed signal, and m represent the length of the receding horizon over which is applied the optimisation algorithm. For details see [11]. We now present an outline of the algorithm as a series of steps:

1. Set $t = 1$
2. Set $(\hat{y}_t, \ldots, \hat{y}_{t+m-1}) = O_{1,m}$ where $O_{1,m}$ denotes a zero matrix of order $1 \times m$
3. Compute the first n lags of the sampled non-central autocovariance of $(\tilde{y}_1, \ldots, \tilde{y}_{t-1}, \hat{y}_t, \hat{y}_{t+1}, \ldots, \hat{y}_{t+m-1})$ via
$$\hat{r}_k := \frac{1}{t+m-1} \sum_{i=k+1}^{t+m-1} \hat{y}_i \hat{y}_{i-k}, \quad k = 0, \ldots, n \tag{14}$$
where we are considering $\hat{y}_i = \tilde{y}_i$ for $i = 1, \ldots, t-1$
4. Generate a new m-tuple $(\hat{y}_t, \ldots, \hat{y}_{t+m+1}) \in \{0,1\}^m$ and repeat step 3 until all m-tuples have been tested.
5. Let $\tilde{y}_t = \hat{y}_t$ for the m-tuple $(\hat{y}_t, \ldots, \hat{y}_{t+m-1}) \in \{0,1\}^m$ for which $\left\| \left\{\hat{r}_i\right\}_{i=0}^{n} - \left\{\tilde{r}_i^d\right\}_{i=0}^{n} \right\|_2$ is minimum.
6. If $t < N$, let $t = t + 1$ and go to step 2.
7. Convert the {0,1} N-tuple $(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_N)$ into a {a,b} N-tuple $(y_1, y_2, \ldots, y_N)$ via
$$y_t := (b-a)\tilde{y}_t + a, \qquad t = 1, \ldots, N \tag{15}$$

For our choice of input signal, we were motivated by the fact that a typical input signal used in system identification is bandlimited white noise, see [12]. For $m = 1$, $N = 1000$ and $n = 50$, we obtain the results presented in Fig. 1. From this figure, we can see that both the autocovariance and spectrum of the generated signal are very similar to those of white noise.

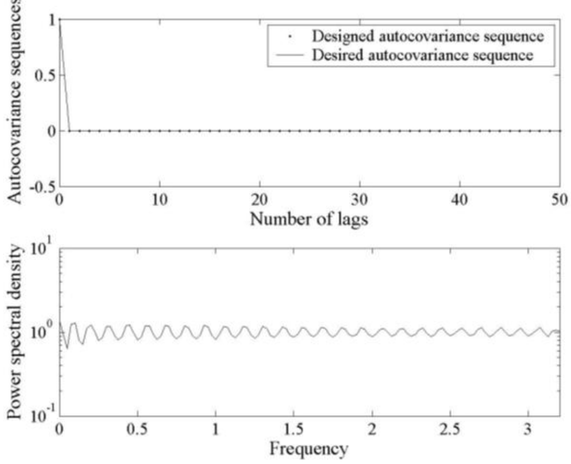

Fig. 1. Characteristics of the generated pseudo white noise signal for m = 1, N = 10^6 and n = 50

IV. SIMULATION RESULTS

To demonstrate the performance of proposed robust procedure for parameters estimation, we consider a single-input/single-output process, described by:

$$y(k) = \frac{B(q^{-1})}{A(q^{-1})} u(k) + \nu(k) \tag{16}$$

(where the delay operator is defined by q^{-1}). $A(q^{-1})$ and $B(q^{-1})$ are polynomials of the second degree, depicted as:

$$\begin{aligned} A(q^{-1}) &= 1 - 1.5q^{-1} + 0.7q^{-2} \\ B(q^{-1}) &= 1q^{-1} + 0.5q^{-2} \end{aligned} \tag{17}$$

From (16), can be seen the dynamics of the process is involved in the noise process. We consider the case when the density for observation noise $\nu(k)$ is Non-Gaussian, with contamination $\varepsilon = 0.1$ and variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 100$, see (10). For excited signal we utilize the generated pseudo white noise, which the amplitude lie in interval $[-1,1]$.

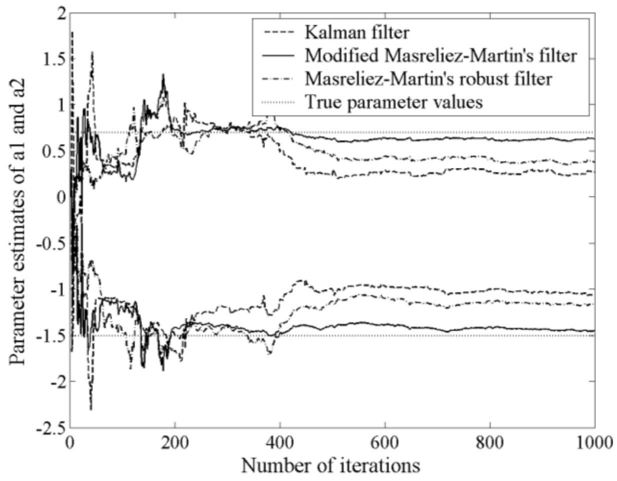

Fig. 2. Estimates of parameters a_1 and a_2

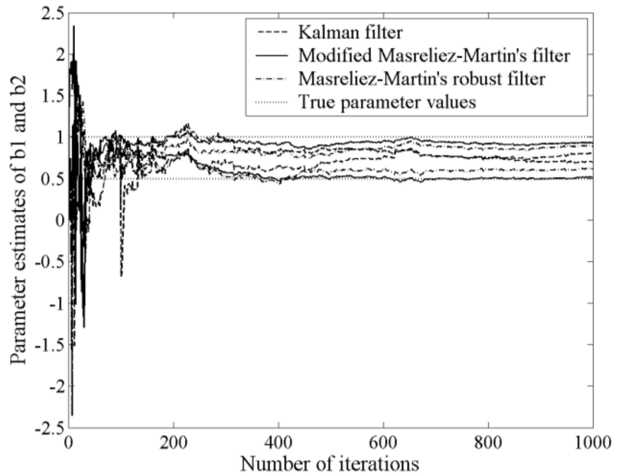

Fig. 3. Estimates of parameters b_1 and b_2

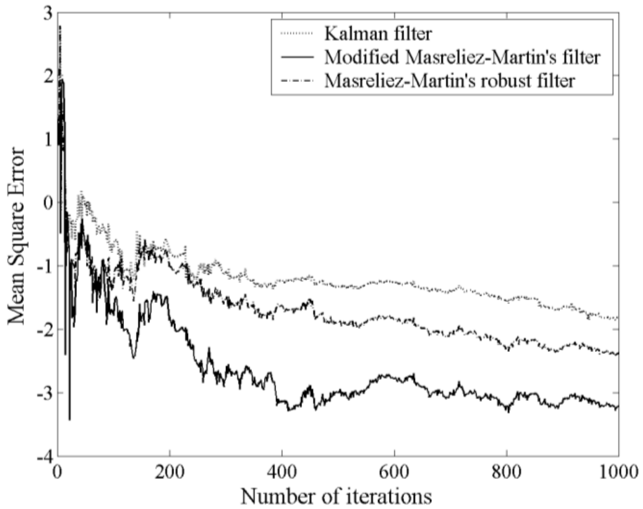

Fig. 4. Mean Square Error

V. CONCLUSION

In this paper we have applied the robust filter theory for solving problem of the robust parameter estimation, in which some modifications of the Masreliez-Martin’s filter are used. Because of its advantages when the plant output measurement is disturbed by noise, output error model is used. As input signal, bandlimited white noise is generated by proposed algorithm based on ideas from predictive control.

The proposed modificated Masreliez-Martin’s filter has satisfactory performances as parameter estimator. However, all the available practically applicable recursive robust estimators, which are obtained as a result of approximations and assumptions, require further practical and/or theoretical verifications.

REFERENCES

[1] I.D. Landau, R. Lozano and M. M'Saad, Adaptive Control, 1st ed., Springer, 1998

[2] L. Dugard and I.D. Landau, “Recursive output error identification algorithms theory and evaluation”, *Automatica*, Vol. 16, Iss. 5, pp. 443-462, 1980

[3] R. E. Kalman, “A new approach to linear filtering and prediction problems”, Transactions of the ASME – Journal of Basic Engineering, 82, pp. 35-45, March 1960.

[4] H. Heffes, “The effect of erroneous models on the Kalman filter response”, *IEEE Transactions on Automatic Control*, AC-11, pp. 541-543, 1966.

[5] T. Nishimura, “On the a priori information in sequential estimation problems”, *IEEE Transactions on Automatic Control*, AC–11, pp. 197-204, 1966.

[6] C. J. Masreliez, “Approximate non-Gaussian filtering with linear state and observation relations”, *IEEE Transactions on Automatic Control*, Vol. 20, Iss. 1, pp. 107–110, 1975.

[7] C. J. Masreliez and R. D. Martin, “Robust Bayesian estimation for the linear model and robustifying the Kalman filter”, *IEEE Transactions on Automatic Control*, Vol. 22, Iss. 3, pp. 361–371, 1977.

[8] C. Tsai and L. Kurz, “An adaptive robustizing approach to Kalman filtering”, *Automatica*, Vol.19, Iss.3, pp. 278-288, May 1983.

[9] P.J. Huber, “Robust estimation of a location parameter”, *The Annals of Mathematical Statistics*, Vol. 35, Num. 1, pp. 73-101, 1964

[10] J. Sacks and D. Ylvisaker, “A note on Huber’s robust estimation of a location parameter”, *The Annals of Mathematical Statistics*, Vol. 43, Num.4, pp. 1068-1075, 1972

[11] R. Rojas, J. Welsh and G. Goodwin, “A receding horizon algorithm to generate binary signals with a prescribed autocovariance”, *American Control Conference*, pp. 122-127, New Jork, 2007

[12] L. Ljung, *System Identification*, 2nd ed., New Jersey: Prentice Hall, 1999.