INFORMATICA, 1991, Vol.2, No.4, 579–592

RECURSIVE ROBUST ESTIMATION OF DYNAMIC SYSTEMS PARAMETERS

Rimantas PUPEIKIS

Institute of Mathematics and Informatics
Lithuanian Academy of Sciences,
2600 Vilnius, Akademijos St.4. Lithuania

Abstract. In the previous paper (Pupeikis, 1990) the problem of model order determination in the presence of outliers in observations has been considered. The aim of the given paper is the development of the recursive algorithms of computation of M-estimates ensuring their stability conditions. In this connection the approach, based on adaptive Huber's monotone psi-function, is worked out. It is also used for the detection of the outliers in time series and for the correction both outliers and M-estimates during successive calculations. The results of numerical simulation by computer (Fig. 1 and Table 1) are given.

Key words: recursive algorithm, outlier, robustness.

Statement of the problem. By identification and parameter estimation of real objects it is often assumed that the additive noise affecting the output of a dynamic system is Gaussian. However in many cases this assumption is not valid because of the outliers in observations, used for parameter estimation. That's why the recursive least squares (RLS) algorithm applied to the current calculation of the unknown parameters appeared to be non-effective. In this case the robust analogues of the RLS, based on the recursive calculations of M-estimates, may be used.

Consider a single input x_k and a single output y_k linear discrete-time system described by the difference equation

$$y_k = -a_1 y_{k-1} - \ldots - a_n y_{k-n} + b_1 x_{k-1} + \ldots + b_n x_{k-n}. \quad (1)$$

Suppose that y_k is observed with an additive noise ξ_k, i.e.,

$$u_k = y_k + \xi_k \quad (2)$$

then

$$\begin{aligned} u_k = &- a_1 u_{k-1} - \ldots - a_n u_{k-n} + b_1 x_{k-1} + \\ &\ldots + b_n x_{k-n} + \xi_k + a_1 \xi_{k-1} + \ldots + a_n \xi_{k-n} \end{aligned} \quad (3)$$

or

$$u_k = \frac{B(z^{-1})}{1 + A(z^{-1})} x_k + \xi_k^* \quad (4)$$

by introducing the backward shift operator z^{-1} defined by $z^{-1} x_k = x_{k-1}$, where

$$\xi_k = (1 - \gamma_k) v_k + \gamma_k \eta_k \quad (5)$$

is the sequence of independent identically distributed variables with ε-contaminated distribution of the shape

$$p(\xi_k) = (1 - \varepsilon) N(0, \sigma_1^2) + \varepsilon N(0, \sigma_2^2), \quad (6)$$

$p(\xi_k)$ is a probability density distribution of the sequence ξ_k; γ_k is a random variable, taking the values of 0 and 1 with the probabilities $p(\gamma_k = 1) = \varepsilon$, $p(\gamma_k = 0) = 1 - \varepsilon$; v_k, η_k are the sequences of independent Gaussian variables with the zero means and σ_1^2, σ_2^2 respectively; n is the order of difference equation (1);

$$\xi_k^* = \left[1 + A(z^{-1})\right]^{-1} \xi_k \quad . \quad (7)$$

is the sequence of the correlated additive noise;

$$B(z^{-1}) = b_1 z^{-1} + \ldots + b_n z^{-n},$$
$$A(z^{-1}) = a_1 z^{-1} + \ldots + a_n z^{-n} \tag{8}$$

are polynomials and

$$a^T = (a_1, \ldots, a_n), \quad b^T = (b_1, \ldots, b_n)$$

are object parameters, subject to estimating.

It is assumed that the roots of $A(z^{-1})$ are outside the unit circle of the z^{-1}-plane. The true orders of the polynomials $A(z^{-1})$ and $B(z^{-1})$ are known. The input signal x_k is persistently exciting of an arbitrary order.

Recursive parameter estimation in the absence of outliers in observations. Suppose that in equation (6) $\varepsilon = 0$, therefore $p(\xi_k) = N(0, \sigma_1^2)$. In this case, as it is shown in (S.Ljung and L.Ljung, 1985) to estimate the vector of the unknown parameters $\theta^T = (a^T, b^T)$ the basic RLS algorithm of the shape

$$\theta_{k+1} = \theta_k + R_{k+1}^{-1} \varphi_{k+1} \epsilon_{k+1}, \tag{9}$$

$$R_{k+1} = \lambda R_k + \varphi_{k+1} \varphi_{k+1}^T, \tag{10}$$

$$\epsilon_{k+1} = u_{k+1} - \varphi_{k+1}^T \theta_k \tag{11}$$

is used, where

$$\theta_{k+1}^T = (\widehat{a}^T, \widehat{b}^T)_{k+1} = (\widehat{a}_1, \ldots, \widehat{a}_n, \widehat{b}_1, \ldots, \widehat{b}_n)_{k+1} \tag{12}$$

is the vector of the unknown parameter estimates after $k+1$ samples;

$$\varphi_{k+1} = (-u_k, \ldots, -u_{k+1-n} x_k, \ldots, x_{k+1-n})^T \tag{13}$$

is the vector of n most recent observations of input x_k and output u_k;

$$\theta_k = R_k^{-1} \sum_{t=1}^{k} \varphi_t u_t \lambda^{k-t}, \tag{14}$$

$$R_k = \sum_{t=1}^{k} \varphi_t \varphi_t^T \lambda^{k-t}; \tag{15}$$

$0.95 \leqslant \lambda \leqslant 1$ is a time-varying weighing factor.

The basic RLS can be rewritten as

$$\theta_{k+1} = \theta_k + K_{k+1} \epsilon_{k+1}, \tag{16}$$

$$K_{k+1} = \frac{P_k \varphi_{k+1}}{\lambda + \varphi_{k+1}^T P_k \varphi_{k+1}}, \tag{17}$$

$$P_{k+1} = \left(P_k - \frac{P_k \varphi_{k+1} \varphi_{k+1}^T P_k}{\lambda + \varphi_{k+1}^T P_k \varphi_{k+1}}\right) \lambda^{-1} \tag{18}$$

$$P_0 = \alpha I, \quad \alpha >> 1 \tag{19}$$

by introducing

$$P_{k+1} = R_{k+1}^{-1} \tag{20}$$

and applying the matrix inversion lemma to (10).

This algorithm minimizes the quadratic loss function

$$V(\theta) = \sum_{t=1}^{s} \lambda^{s-t} \epsilon_t^2 + V_0 \tag{21}$$

if λ is constant and

$$V(\theta) = \sum_{t=1}^{s} \left(\prod_{j=t}^{s} \lambda_j\right) \epsilon_t^2 + V_0 \tag{22}$$

in the c.posite case (Rao Sripada and Grant Fisher, 1987).

In equations (21),(22)

$$V_0 = (\theta - \theta_0)^T P_0^{-1} (\theta - \theta_0) \tag{23}$$

and depends on the initial conditions of RLS.

It is known (Åström and Eykhoff, 1971) that under the above mentioned and some other conditions RLS is going to have the maximal convergence rate.

Recursive parameter estimation in the presence of outliers in observations. It was assumed earlier that in equation (6) $\varepsilon = 0$. Now let us consider such a case, when this assumption is invalid. Then RLS used for the current estimation of unknown parameters of a mathematical model of the dynamic object (1)–(8) becomes of a little use. As a result the computer time has been uselessly wasted to obtain unsatisfactory results. In this case instead RLS the algorithms of computation of M-estimates, which are worked out by (Novovičova, 1987) may be used.

These algorithms are:

$$\theta_{k+1} = \theta_k + \frac{P_k \varphi_{k+1} \widehat{\sigma} \psi(r_{k+1}^{(k)}/\widehat{\sigma})}{\left[\psi'(r_{k+1}^{(k)}/\widehat{\sigma})\right]^{-1} + \varphi_{k+1}^T P_k \varphi_{k+1}}, \tag{24}$$

$$P_{k+1} = P_k - \frac{P_k \varphi_{k+1} \varphi_{k+1}^T P_k}{\left[\psi'(r_{k+1}^{(k)}/\widehat{\sigma})\right]^{-1} + \varphi_{k+1}^T P_k \varphi_{k+1}} \tag{25}$$

S-algorithm,

$$\theta_{k+1} = \theta_k + \frac{P_k \varphi_{k+1} \widehat{\sigma} \psi(r_{k+1}^{(k)}/\widehat{\sigma})}{1 + \varphi_{k+1}^T P_k \varphi_{k+1}}, \tag{26}$$

$$P_{k+1} = P_k - \frac{P_k \varphi_{k+1} \varphi_{k+1}^T P_k}{1 + \varphi_{k+1}^T P_k \varphi_{k+1}} \tag{27}$$

H-algorithm,

$$\theta_{k+1} = \theta_k + \frac{P_k \varphi_{k+1} \widehat{\sigma} r_{k+1}^{(k)}}{[w_{k+1}^{(k)}]^{-1} + \varphi_{k+1}^T P_k \varphi_{k+1}}, \qquad (28)$$

$$P_{k+1} = P_k - \frac{P_k \varphi_{k+1} \varphi_{k+1}^T P_k}{[w_{k+1}^{(k)}]^{-1} + \varphi_{k+1}^T P_k \varphi_{k+1}}, \qquad (29)$$

$$w_{k+1}^{(k)} = \begin{cases} \widehat{\sigma} \psi(r_{k+1}^{(k)}/\widehat{\sigma})/r_{k+1}^{(k)} & \text{for } r_{k+1}^{(k)} \neq 0 \\ \rho_o'' & \text{for } r_{k+1}^{(k)} = 0, \end{cases} \qquad (30)$$

$$r_{k+1}^{(k)} = u_{k+1} - \varphi_{k+1}^T \theta_k \qquad (31)$$

and W-algorithm, generating current M-estimates by means of minimizing sums

$$\sum_{i=1}^{p} \rho\left(\frac{e_i}{\sigma}\right) = \min \qquad (32)$$

or by solving the system of nonlinear equations

$$\sum_{i=1}^{p} \psi\left(\frac{e_i}{\sigma}\right) \varphi_i = 0 \qquad (33)$$

if the derivatives with respect to θ are taken. Here $\rho(\cdot)$ is a symmetric robustifying loss function, $\psi = \rho'(\cdot)$, $\widehat{\sigma}$ denotes an estimate of the innovations scale and may be obtained simultaneously (Novovičova, 1987).

As initial values for the above mentioned algorithms the least square estimates obtained for small data set can be used.

Adaptive Huber's psi-function. There have been worked out various psi-functions for M-estimators (Stockinger and Dutter,1987). The most popular of them is a Huber's psi-function which can be written as

$$\psi_H(t) = \begin{cases} t & \text{if} \quad |t| \leqslant c_H \\ c_H \operatorname{sign} t & \text{if} \quad |t| > c_H, \end{cases} \tag{34}$$

where $c_H > 0$ is given, $t = r_{k+1}^{(k)}/\widehat{\sigma}$.

Unfortunately, the problem of the choice of c_H while estimating the parameters has not been solved by now. That's why therefore appear two situations – the observations will be damaged if we choose c_H too small or the outliers will be let pass without a special processing in the opposite case. Besides, current estimates $a_{k+1}^T = (\widehat{a}_1, \ldots, \widehat{a}_n)_{k+1}$ may be obtained outside the permissible stability area Ω of the parameters of a respective difference equation. In this connection there appears the problem to choose such an adaptive c_H which will ensure the current estimates of the parameters $a^T = (a_1, \ldots, a_n)$ inside of the mentioned area. Therefore it will be shown lower how to choose c_H for the simple second order system.

Suppose, that the problem of recursive computation of M-estimates of the dynamic system described by the equation

$$u_k + a_1 u_{k-1} + a_2 u_{k-2} = b_1 x_{k-1} + b_2 x_{k-2} + \xi_k \tag{35}$$

with

$$\Omega = \{a : a_1 - a_2 < 1, \quad a_1 + a_2 > -1, \quad -1 < a_2 < 1\} \tag{36}$$

is solved.

Then four inequalities take place:

$$\begin{aligned} &\widehat{a}_{1_{k+1}} + \widehat{a}_{2_{k+1}} = \widehat{a}_{1_k} + \widehat{a}_{2_k} + \omega_{k+1}\, \widehat{\sigma}_{k+1}\, \psi(t) > -1, \\ &\widehat{a}_{1_{k+1}} - \widehat{a}_{2_{k+1}} = \widehat{a}_{1_k} - \widehat{a}_{2_k} + \nu_{k+1}\, \widehat{\sigma}_{k+1}\, \psi(t) < 1, \\ &\widehat{a}_{2_{k+1}} = \widehat{a}_{2_k} + \beta_{k+1}\, \widehat{\sigma}_{k+1}\, \psi(t) < 1, \\ &\widehat{a}_{2_{k+1}} = \widehat{a}_{2_k} + \beta_{k+1}\, \widehat{\sigma}_{k+1}\, \psi(t) > -1, \end{aligned} \tag{37}$$

which can be rewritten as

$$-\omega_{k+1}\,\widehat{\sigma}_{k+1}\,\psi(t) < 1 + \widehat{a}_{1_k} + \widehat{a}_{2_k},$$
$$\nu_{k+1}\,\widehat{\sigma}_{k+1}\,\psi(t) < 1 - \widehat{a}_{1_k} + \widehat{a}_{2_k},$$
$$-1 - \widehat{a}_{2_k} < \beta_{k+1}\,\widehat{\sigma}_{k+1}\,\psi(t) < 1 - \widehat{a}_{2_k},$$

where

$$\omega_{k+1} = \gamma_{k+1} + \beta_{k+1},$$
$$\nu_{k+1} = \gamma_{k+1} - \beta_{k+1},$$
$$\gamma_{k+1} = \frac{P_k(1,1)\varphi_{k+1}(1) + P_k(1,2)\varphi_{k+1}(2)}{1 + \varphi_{k+1}^T P_k \varphi_{k+1}}$$
$$\beta_{k+1} = \frac{P_k(2,1)\varphi_{k+1}(1) + P_k(2,2)\varphi_{k+1}(2)}{1 + \varphi_{k+1}^T P_k \varphi_{k+1}}$$

if H-algorithm is used, $P_k(i,j)$ are the respective elements of the matrix P_k.

Thus,

$$c_H = \min\{t_{1_k},\ t_{2_k},\ t_{3_k}\}, \tag{38}$$

where

$$t_{1_k} = \frac{1 + \widehat{a}_{1_k} + \widehat{a}_{2_k}}{-\omega_{k+1}\,\widehat{c}_{k+1}},$$
$$t_{2_k} = \frac{1 - \widehat{a}_{1_k} + \widehat{a}_{2_k}}{\nu_{k+1}\,\widehat{\sigma}_{k+1}},$$
$$t_{3_k} = \begin{cases} \frac{1-\widehat{a}_{2_k}}{\beta_{k+1}\widehat{\sigma}_{k+1}} & \text{if} \quad \beta_{k+1} > 0 \\ \frac{1+\widehat{a}_{2_k}}{\beta_{k+1}\widehat{\sigma}_{k+1}} & \text{if} \quad \beta_{k+1} < 0. \end{cases}$$

It is known that for great order systems the parameter stability inequalities are non-linear. In this case a decomposition of the initial system into a successive combination of simple systems may be used.

Notice that for first and second order systems some simulation results obtained by using the adaptive psi-function is given in (Pupeikis, 1989; Kazlauskas and Pupeikis, 1991).

Outliers correction. It is obvious that the above developed approach, based on the adaptive Huber's psi-function, may be also used for the detection and correction of outliers in observations. Suppose that in the k-th time of recursive calculations there appears an outlier u_{k+1}. Then according to (34) for second order system (35) psi-function becomes

$$\psi_H(t) = t \tag{39}$$

if all four inequalities are satisfied or

$$\psi_H(t) = c_H \text{ sign } t \tag{40}$$

if at least one of them is invalid, where c_H is of the shape (38).

In order to decrease the outliers influence on the accuracy of the next calculations. a correction may be done in the following way:

$$u_{k+1} = \widehat{\sigma}_k \, c_H + \varphi_{k+1}^T \, \theta_k. \tag{41}$$

The case, when there exists outlier, which could not be detected using inequalities (37), is very special and more sensitive methods for the outliers detection and the following correction are required.

Simulation results. The efficiency of H-algorithm in the presence of outliers was investigated by numerical simulation by means of a computer. The noiseless sequence y_k was generated by the equation from the paper (Åström and Eykhoff, 1971)

$$y_k = \frac{z^{-1} + 0.5z^{-2}}{1 - 1.5z^{-1} + 0.7z^{-2}} x_k \qquad (k = \overline{1,500}). \tag{42}$$

In the capacity of the input x_k the realizations of the sequences of independent Gaussian variables ζ_k with zero mean and unitary dispersion and of first order AR model of the shape

$$x_k = 0.9\, x_{k-1} + 0.43\, \zeta_k \tag{43}$$

were used. As the additive noise ξ_k^* the realization of the discrete AR process was generated according to the equation

$$\xi_k^* = (1 - 1.5\, z^{-1} + 0.7 z^{-2})^{-1} \xi_k, \tag{44}$$

where ξ_k is a sequence of independent identically distributed variables of shape (5) with ε – contaminated distribution of shape (6) and $\sigma_1^2 = 1$, $\sigma_2^2 = 100$. 10 experiments with different realizations of the noise ξ_k^* at the noise level $\sigma_{\xi^*}^2 / \sigma_y^2 = 0.5$ were carried out. In each i-th experiment the initial P_0, θ_0 for H-algorithm with constant and adaptive c_H were obtained using formulas:

$$P_0 = \left\{ \sum_{j=4}^{50} \varphi_j \varphi_j^T \right\}^{-1},$$

$$\theta_0 = -P_0 \sum_{j=4}^{50} \varphi_j u_j.$$

While simulating was assumed that in expression (6) $\varepsilon = 0.5$ and in Huber's psi-function (34) $c_H = 0.5$. The current M-estimates of the vector θ by means of the H-algorithm with classical Huber's psi-function (34) and adaptive one (34), (38) were received.

In Fig.1 the averaged by 10 experiments square parameter errors related to square true parameters according to

$$W = \frac{1}{10} \sum_{i=1}^{10} W^{(i)}, \tag{45}$$

for an object (42), (44) are presented. Here

$$W^{(i)} = \frac{\sum_{j=1}^{4}(\theta_{j_k}^{(i)} - \theta_j)^2}{\sum_{j=1}^{4} \theta_j^2},$$

where

$$\theta_{1_k}^{(i)} = \widehat{a}_{1_k}^{(i)},\ \theta_{2_k}^{(i)} = \widehat{a}_{2_k}^{(i)},\ \theta_{3_k}^{(i)} = \widehat{b}_{1_k}^{(i)},\ \theta_{4_k}^{(i)} = \widehat{b}_{2_k}^{(i)},$$

are the estimates of the respective parameters of the equation (42) on i-th experiment and k-th time,

$$\theta_1 = a_1,\ \theta_2 = a_2,\ \theta_3 = b_1,\ \theta_4 = b_2$$

are the true parameters of the mentioned equation.

In Table 1 the averaged by 10 experiments variables (45) and their confidence intervals, obtained using classical formulas (Bendat and Piersol, 1971) and calculated for different inputs are given. In this connection the first line of each input

Table 1. Averaged values (45) and their confidence intervals depending on observations

50	200	350	500
Input – Gaussian process			
16.047±4.516	9.109±3.191	7.116±2.55	6.1±2.362
15.688±4.391	8.713±2.056	5.586±1.72	4.49±1.503
Input – AR process			
70.553±23.359	49.592±20.91	42.386±19.517	38.302±18.108
70.781±26.327	37.824±13.965	24±12.754	20.718±9.48

corresponds to the values (45), which were calculated by means of H-algorithm with constant c_H. The second line corresponds to the same values which were obtained using H-algorithm and adaptive c_H. From the simulation results, presented in Fig.1 and Table 1, it follows, that for the different inputs the accuracy of the M-estimates will be different too independently of choosing c_H in (34). On the other hand accuracy of the M-estimates for the same input depends on choosing of c_H. If we calculate it, using the approach presented here , we will reach a higher accuracy comparing with constant c_H.

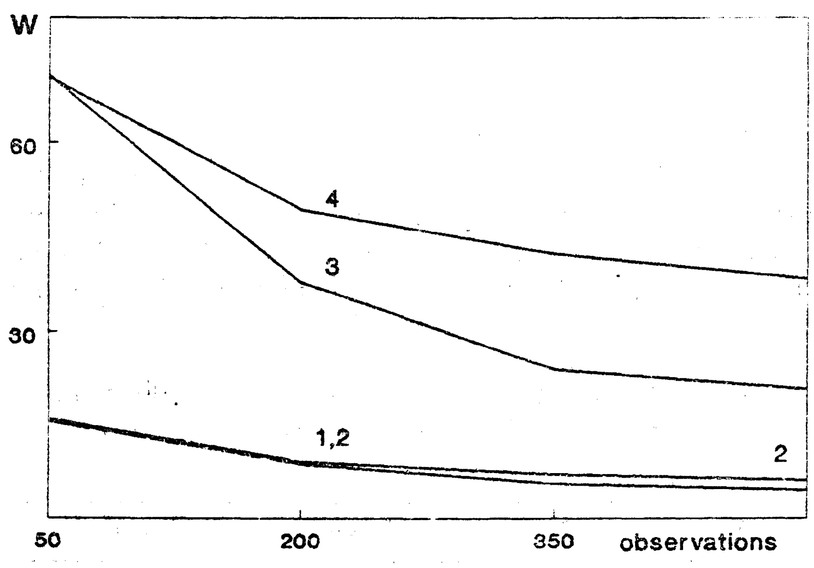

Fig. 1. Averaged values (45) depending on observations. Input: Gaussian process – curve 1, 2; AR process – 3, 4. c_H: constant – curve 2, 4; adaptive – 1, 3.

Conclusions. For the determination of c_H in Huber's psi-function during recursive parameter estimation it is possible to use the above presented approach based on checking of the stability conditions of the difference equation parameters. The results of numerical simulation, carried out by computer prove the usefulness of the proposed approach.

REFERENCES

Åström, K.J., and P.Eykhoff (1971). System identification – a survey. *Automatica*, **7**(2), 123–162 .

Bendat, J.S., and A.G.Piersol (1971). *Measurement and Analysis of Random Data*. Mir, Moscow. 408pp. (in Russian).

Kazlauskas, K., and R.Pupeikis (1991). *Data Processing Digital Systems*. Mokslas, Vilnius. 222pp. (in Russian).

Ljung, S., and L.Ljung (1985). Error propagation properties of recursive least-squares adaptation algorithms. *Automatica*, **21**(2), 157–167.

Novovičova, J. (1987). Recursive computation of M-estimates for the parameters of the linear dynamical system. *Problems of Control and Information Theory*, **16**(1), 19–59.

Pupeikis, R. (1989). Optimal recursive identification methods (4. Algorithms adaptive stabilization in case of outliers). *Works of the Academy of Sciences the Lithuanian SSR*, Ser. B. **5**(174), 134–142 (in Russian).

Pupeikis, R. (1990). Model order robust determination. *Informatica*, **1**(2), 96–109.

Rao Sripada, N., and D.Grant Fisher (1987). Improved least squares identification. *Int.J.Control*, **46**(6), 1889–1913.

Stockinger, N., and R.Dutter (1987). Robust time-series analysis, a overview. *Kybernetika*, **23**(1–5), 90pp.

Received May 1991

R. Pupeikis received the Degree of Candidate of Technical Sciences from the Kaunas Polytechnic Institute, Kaunas, Lithuania, 1979. He is a senior research worker of the Department of the Technological Process Control at the Institute of Mathematics and Informatics. His research interests include the classical and robust approaches of dynamic system identification as well technological process control.