INFORMATICA, 1998, Vol. 9, No. 3, 325–342

State Estimation of Dynamic Systems in the Presence of Time-Varying Outliers in Observations

Rimantas PUPEIKIS
Institute of Mathematics and Informatics
Akademijos 4, 2600 Vilnius, Lithuania
e-mail: pupeikis@ktl.mii.lt

Received: March 1998

Abstract. In the previous papers (Masreliez and Martin, 1977; Novovičova, 1987; Schick and Mitter, 1994) the problem of recursive estimation of linear dynamic systems parameters and of the state of such systems in the presence of outliers in observations have been considered. In this connection various ordinary recursive techniques are worked out, when systems output is corrupted by an additive noise with a time homogeneous contamination of outliers. The aim of the given paper is the development of an approach for robust recursive state estimation of linear dynamic systems in a case of additive noises with time-varying outliers. The recursive technique based on the abovementioned theoretical results is obtained and proved by state estimation of the real chemical process (Box and Jenkins, 1970). The results of numerical simulation by computer (Fig. 1–3) are given.

Key words: dynamic system, Kalman filter, robustness, state estimation, optimization.

1. Introduction

The Kalman optimal recursive filter applied in state estimation appeared to be inefficient, in the presence of outliers in observations (Masreliez and Martin, 1977). That is why multivariate recursive robust approaches and algorithms are worked out (Schick and Mitter, 1994). On the other hand, the theoretical ground of those alternatives is based on the classical robust theory of the estimate of a location parameter using the stochastic models with time-homogeneous contamination of outliers (Huber, 1964). It is known, that the accuracy of estimates, which are invariant for the permutation of observations, doesn't depend on the disposition of outliers in observations to be processed. However, referring to the dynamic discrete-time processes, the disposition of outliers turns out to be very important. If various recursive robust algorithms some times turns out to be efficient in the presence of rare and isolated outliers, but not when the outliers occur in batches (Schick and Mitter, 1994), then there always arise special problems in the presence of time-varying outliers in observations (Pupeikis and Huber, 1997). In this case it is important to solve the generalized problem of a model of outliers, which are varying in time, including, as an extreme case, the patchy outliers. A new approach to be proposed here is based on the robust filtering by means of a bank of parallel Kalman filters, at each time moment generating, correspondingly, a bank of state estimates, and in the procedure of optimization of the state estimation itself, choosing, at each time moment, the optimal current estimate with the current minimal filtering error. It is theoretically shown here that a vector of state estimates, guaranteeing the minimal variance of reconstructed input, also guarantees a minimal square filtering error, which is actually unknown. That is why it is possible to construct an optimization procedure in order to choose from a bank of state estimates at each time moment such an estimate, which would be the optimal one in the sense of a minimal square filtering error.

2. Statement of the Problem

Assume that we consider a single input $\boldsymbol{\mu}_k$ and single output $\mathbf{x}_k$ of a linear discrete-time system, described by the difference equation

$$x_k = \mu_k + a_1 x_{k-1} + \ldots + a_n x_{k-n}, \tag{1}$$

and that $\mathbf{x}_k$ is observed under additive noise $\mathbf{Z}_k$, i.e.,

$$u_k = x_k + z_k, \tag{2}$$

where

$$x_k = W(q^{-1}; \mathbf{a})\mu_k \tag{3}$$

is the value of unobserved output at a time moment k, and

$$W(q^{-1}; \mathbf{a}) = \frac{1}{1 - A(q^{-1}; \mathbf{a})} \tag{4}$$

is a system transfer function,

$$A(q^{-1}; \mathbf{a}) = \sum_{i=1}^{n} a_i q^{-i}, \tag{5}$$

$$\mathbf{a}^T = (a_1, \ldots, a_n), \tag{6}$$

$a_1, \ldots, a_n$ are parameters of a polynomial (5), q^{-1} is the backward shift operator defined by

$$x_{k-n} = x_k q^{-n}, \tag{7}$$

u_k is the observed value of output; $\mu_{k-1}, \ldots, \mu_{k-n}$, and $x_{k-1}, \ldots, x_{k-n}$ are unobserved values of input and output, respectively; $\boldsymbol{\mu}_k \sim \mathcal{N}(0, \sigma_\mu^2)$ is a sequence of independent identically distributed variables; $\mathbf{Z}_k$ is a sequence of independent identically distributed variables with an 'ε – contaminated' distribution of the form

$$p(z_k) = (1 - \varepsilon_k)\mathcal{N}(0, \sigma_\xi^2) + \varepsilon_k \mathcal{N}(0, \sigma_\nu^2) \tag{8}$$

and the variance

$$\sigma_z^2 = (1 - \varepsilon_k)\sigma_\xi^2 + \varepsilon_k \sigma_\nu^2; \tag{9}$$

$p(z_k)$ is a probability density distribution of the sequence $\mathbf{Z}_k$;

$$z_k = (1 - \gamma_k)\xi_k + \gamma_k \nu_k \tag{10}$$

is the value of additive noise at time moment k; γ_k is a random variable, taking values 0 or 1 with probabilities $p(\gamma_k = 0) = 1 - \varepsilon_k$, $p(\gamma_k = 1) = \varepsilon_k$; $\xi_k,\ \nu_k$ are sequences of independent Gaussian variables with zero means and variances $\sigma_\xi^2,\ \sigma_\nu^2$, respectively; besides, $\sigma_\xi < \sigma_\nu$; $0 \leqslant \varepsilon_k \leqslant 1$ is the unknown fraction of 'contamination' varying in time.

It is supposed that the roots of $A(q^{-1}; \mathbf{a})$ are outside the unit circle of the q^{-1}. The true order n of the polynomial $A(q^{-1}; \mathbf{a})$ and true values of the parameters $a_1, \ldots, a_n$ are known. The input signal $\boldsymbol{\mu}_k$ corresponds to the persistent exitation conditions of arbitrary order according to Eykhoff (1974).

The aim of the given paper is the development of an approach for a robust recursive estimation of states $x_k, x_{k-1}, \ldots, x_{k-n}$ of a linear dynamic system (1), (2) in the presence of time-varying outliers (8)–(10) in observations $u_1, u_2, \ldots, u_N$ of an output $\mathbf{U}_k$.

3. Robustifying the Kalman Filter

Linear discrete-time system (1), (2) can be described using the state equations of the form

$$\boldsymbol{\beta}(k+1) = \mathbf{A}\boldsymbol{\beta}(k) + \mathbf{h}\mu_{k+1}, \tag{11}$$
$$u_k = \mathbf{c}^T \boldsymbol{\beta}(k) + z_k, \tag{12}$$

where

$$\boldsymbol{\beta}(k+1) = (\beta_1(k+1), \beta_2(k+1), \ldots, \beta_n(k+1))^T, \tag{13}$$
$$\beta_1(k+1) = x_{k+1}, \beta_2(k+1) = x_k, \ldots, \beta_n(k+1) = x_{k-n+2}, \tag{14}$$
$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \\ 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 1 & 0 \end{bmatrix}, \tag{15}$$
$$\boldsymbol{\beta}(k) = (\beta_1(k), \beta_2(k), \ldots, \beta_n(k))^T,$$

$$\beta_1(k) = x_k, \beta_2(k) = x_{k-1}, \ldots, \beta_n(k) = x_{k-n+1}, \tag{16}$$
$$\mathbf{h}^T = \mathbf{c}^T = [1\ 0\ \ldots 0]. \tag{17}$$

It is known (Meyr and Spies, 1984), that it is possible to reduce the influence of outliers on the accuracy of estimates by discarding the observations, after their processing yielding relatively 'too large' residuals. In this case, for the state-estimation of linear dynamic system (11), (12) the robust Kalman filter

$$\hat{\boldsymbol{\beta}}(k+1) = A\hat{\boldsymbol{\beta}}(k) + \mathbf{k}(k+1)\psi(e_{k+1}), \tag{18}$$
$$\mathbf{k}(k+1) = \mathbf{Q}(k+1)\mathbf{c}[\sigma_z^2 + \mathbf{c}^T\mathbf{Q}(k+1)\mathbf{c}]^{-1}, \tag{19}$$
$$\mathbf{Q}(k+1) = \mathbf{A}\mathbf{P}(k)\mathbf{A}^T + \sigma_\mu^2, \tag{20}$$
$$\mathbf{P}(k) = \mathbf{Q}(k) - \mathbf{k}(k)\mathbf{c}^T\mathbf{Q}(k) \tag{21}$$

can be used, where

$$e_{k+1} = u_{k+1} - \mathbf{c}^T\mathbf{A}\hat{\boldsymbol{\beta}}(k) \tag{22}$$

is an innovation at a time moment $k+1$;

$$\psi(e_{k+1}) = \begin{cases} -\Delta & \text{if} \quad e_{k+1} < -\Delta, \\ e_{k+1} & \text{if} \quad -\Delta \leqslant e_{k+1} \leqslant \Delta, \\ \Delta & \text{if} \quad e_{k+1} > \Delta \end{cases} \tag{23}$$

is Hubers' ψ function, $\hat{\boldsymbol{\beta}}(k)$ is the estimate of $\boldsymbol{\beta}(k)$ at a time moment *k*.

The threshold Δ in (23) depends on the standard deviation σ_ξ of the ground-distribution (normal distribution) and on a fraction of a 'contamination' ε. In general case, σ_ξ is a priori unknown. Drewelow, (1990) proposed to estimate the value of σ_ξ according to the iterative formula

$$\hat{\sigma}_\xi^2(i+1) = \frac{\tau}{s}\sum_{j=1}^{s}\psi\left(\frac{e_j}{\hat{\sigma}_\xi(i)}\right)^2,$$

where τ is a constant,

$$\psi(t) = \begin{cases} -\Delta & \text{if} \quad t < -\Delta, \\ t & \text{if} \quad -\Delta \leqslant t \leqslant \Delta, \\ \Delta & \text{if} \quad t > \Delta, \end{cases}$$
$$t = \frac{e_j}{\hat{\sigma}_\xi(i)};$$

$\hat{\sigma}_\xi(i)$ is the estimate of σ_ξ at the i-th iteration; s is a sample size; i is the i-th iteration of calculations.

There also exist other propositions for determination of a threshold (Hampel *et al.*, 1986; Ljung, 1991; Verboon, 1994). The Huber monotone ψ–function or other functions

(Huber, 1981; Hampel *et al.*, 1986; Stockinger and Dutter, 1987) may be used as a ψ–function.

4. Determination of Criteria for Optimizing the State Estimation

In order to select a function to be minimized in the problem of optimization of the state estimation it is important to determine a relation between the of used but unknown filtering error and the characteristics, which are known beforehand or could be easily calculated. Further such a relation between the filtering error (to be more precise, the averaged square error of prediction of the state) and the variance of reconstructed input $\boldsymbol{\mu}_k$ will be used.

Suppose, that N observations $u_1, u_2, \ldots, u_N$ of an output $\mathbf{U}_k$ are obtained. Then the bank of state estimates $\hat{x}_1(1), \ldots, \hat{x}_N(1); \hat{x}_1(2), \ldots, \hat{x}_N(2); \ldots; \hat{x}_1(L), \ldots, \hat{x}_N(L)$ is calculated by processing $u_1, u_2, \ldots u_N$ using the bank of robust parallel L Kalman filters

$$\hat{\boldsymbol{\beta}}_i(k+1) = \mathbf{A}\hat{\boldsymbol{\beta}}_i(k) + \mathbf{k}(k+1)\psi_i(u_{k+1} - \mathbf{c}^T\mathbf{A}\hat{\boldsymbol{\beta}}_i(k)) \quad \text{for} \quad i = 1, 2, \ldots, L. \tag{24}$$

Here

$$\begin{aligned}\hat{\boldsymbol{\beta}}_i(k+1) &= (\hat{\beta}_1(k+1), \hat{\beta}_2(k+1), \ldots \hat{\beta}_n(k+1))_i^T \\ &= (\hat{x}_{k+1}(i), \hat{x}_k(i), \ldots, \hat{x}_{k-n+2}(i))^T \quad \text{for} \quad i = 1, 2, \ldots, L,\end{aligned} \tag{25}$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_i(k) &= (\hat{\beta}_1(k), \hat{\beta}_2(k), \ldots, \hat{\beta}_n(k))_i^T \\ &= (\hat{x}_k(i), \hat{x}_{k-1}(i), \ldots, \hat{x}_{k-n+1}(i))^T \quad \text{for} \quad i = 1, 2, \ldots, L,\end{aligned} \tag{26}$$

$$\begin{aligned}&\psi_i(u_{k+1} - \mathbf{c}^T\mathbf{A}\hat{\boldsymbol{\beta}}_i(k)) = \psi_i(e_{k+1}(i)) \\ &= \begin{cases} -\Delta_i & \text{if} \quad e_{k+1}(i) < -\Delta_i, \\ e_{k+1}(i) & \text{if} \quad -\Delta_i \leqslant e_{k+1}(i) \leqslant \Delta_i, \\ \Delta_i & \text{if} \quad e_{k+1}(i) > \Delta_i, \end{cases} \quad \text{for} \quad i = 1, 2, \ldots, L.\end{aligned} \tag{27}$$

It can be mentioned that at a fixed time moment k a gain $\mathbf{k}(k)$ is obtained using formula (19) and is the same for all the L filters. The filters in bank (24) are different because of the threshold $\Delta_i \; \forall \, i = 1, 2, \ldots, L$ in (27) besides $\Delta_1 < \Delta_2 < \ldots < \Delta_L$.

Lemma 1. *Assume that in the bank of L robust Kalman filters* (24)–(27) *there exists such a l-th filter, which could guarantee that the filtering error*

$$\widetilde{e}(l) = N^{-1}\mathbf{w}^T\mathbf{w} \tag{28}$$

is 0 for large enough N, i.e.,

$$\hat{x}_k(l) = x_k \qquad \text{for} \quad k = 1, 2, \ldots, N. \tag{29}$$

Then the variance $\sigma^2_{\mu(l)}$ *of the process* $\boldsymbol{\mu}_k(l)$ *for*

$$\mu_k(l) = \hat{x}_k(l) - a_1\hat{x}_{k-1}(l) - \ldots - a_n\hat{x}_{k-n}(l) \quad \textit{for} \quad k = 1, 2, \ldots, N \tag{30}$$

is a minimal and equal to σ^2_μ*, if the initial conditions for* $\boldsymbol{\mu}_k(l)$ *and* $\boldsymbol{\mu}_k$ *are equal, too. Here*

$$\mathbf{w} = (x_1 - \hat{x}_1(l), \ldots, x_n - \hat{x}_N(l))^T \tag{31}$$

is a vector of filtering errors.

Proof. Assume that

$$\widetilde{e}(l) \neq 0, \tag{32}$$

for any l, then

$$\hat{x}_k(l) = x_k + \xi_k(l) \quad \text{for } k = \overline{1, N}, \tag{33}$$

where $\xi_k(l) \sim \mathcal{N}(0, \sigma^2_{\xi(l)})$, besides, $\sigma^2_{\xi(l)} < \sigma^2_z$and $\sigma^2_z$is of the form (9).

The filtering error (32) can be rewritten as

$$\widetilde{e}(l) = (1 - N^{-1})\sigma^2_{\xi(l)}. \tag{34}$$

Then the variance of the process $\boldsymbol{\mu}_k(l)$ is

$$\sigma^2_{\mu(l)} = \sigma^2_\mu + \sigma^2_{\theta(l)} > \sigma^2_\mu, \tag{35}$$

if in (30) $\hat{x}_k(l)$ is replaced by $x_k + \xi_k(l)$ for $k = \overline{1, N}$.

Here $\sigma^2_{\theta(l)}$ is the variance of the process $\theta(l)$, where

$$\theta_k(l) = W^{-1}(q^{-1}; \mathbf{a})\xi_k(l) \quad k = 1, 2, \ldots, N$$

is its value at a time moment k; $W^{-1}(q^{-1}; \mathbf{a})$ is a discrete-time transfer function, inverse to transfer function (4) of the process $\mathbf{x}_k$.

REMARK 1. From (35) we get, that $\sigma^2_{\mu(l)}$ is minimal and equal to σ^2_μ, if $\sigma^2_{\theta(l)}$ is equal to zero. It follows that Lemma 1 gives us the lower bound for $\sigma^2_{\mu(l)}$, $i = 1, 2, \ldots, L$.

Lemma 2. *Let us assume in the bank of* L *robust Kalman filters* (24)–(27) *there exists such a* j*-th filter, which could guarantee that*

$$\hat{x}_k(j) = u_k \quad \textit{for} \quad k = 1, 2, \ldots, N. \tag{36}$$

Then the filtering error

$$\widetilde{e}(j) = N^{-1}\widetilde{\mathbf{w}}^T\widetilde{\mathbf{w}}, \tag{37}$$

and the variance $\sigma^2_{\mu(j)}$ *of the process* $\boldsymbol{\mu}_k(j)$ *for*

$$\mu_k(j) = \hat{x}_k(j) - a_1\hat{x}_{k-1}(j) - \ldots - a_n\hat{x}_{k-n}(j) \tag{38}$$

acquires the maximum.

Here

$$\widetilde{\mathbf{w}} = (x_1 - \hat{x}_1(j), \ldots, x_N - \hat{x}_N(j))^T \tag{39}$$

is a vector of filtering errors.

Proof. It follows from (36) that

$$\widetilde{e}(j) > \widetilde{e}(l) \tag{40}$$

and

$$\sigma^2_{\mu(j)} = \sigma^2_\mu + \sigma^2_b > \sigma^2_{\mu(l)}, \tag{41}$$

for any l, where $\widetilde{e}(l)$ and $\sigma^2_{\mu(l)}$ are defined by (34) and (35), respectively; σ^2_b is the variance of the process $\mathbf{b}_k$ with

$$b_k = W^{-1}(q^{-1}; \mathbf{a})z_k. \tag{42}$$

Obviously, $\sigma^2_{\mu(j)} = \sigma^2_\mu$ in (41), if $\hat{x}_k(j) = x_k$ for $k = \overline{1, N}$ only.

REMARK 2. Lemma 2 gives us the upper bound for $\sigma^2_{\mu(i)}$ $i = 1, 2, \ldots, L$.

PROPOSITION. Suppose that in the bank of L robust Kalman filters (24)–(27) there exists such a f-th filter, which could guarantee that

$$\hat{x}_k(f) = x_k + \xi_k(f) \quad \text{for} \quad k = \overline{1, N}, \tag{43}$$

where $\xi_k(f) \sim \mathcal{N}(0, \sigma^2_{\xi(f)})$ and $\sigma^2_{\xi(f)} < \sigma^2_{\xi(\nu)}$ for any $\nu = 1, 2, \ldots, L-1$. Then after processing N observations this filter has the minimal filtering error

$$\widetilde{e}(f) = (1 - N^{-1})\sigma^2_{\xi(f)}, \tag{44}$$

and the minimal variance

$$\sigma^2_{\mu(f)} = \sigma^2_\mu + \sigma^2_{\theta(f)}. \tag{45}$$

Here $\sigma^2_{\theta(f)}$ is the variance of the process $\theta(f)$ as

$$\theta_k(f) = W^{-1}(q^{-1}; \mathbf{a})\xi_k(f). \tag{46}$$

REMARK 3. Proposition follows from Lemma 1 and Lemma 2.

Proposition lets us to formulate the Theorem in order to choose criteria for the optimization of state estimation.

Theorem 1. *The functions*

$$Q_i(\mathbf{x}, \hat{\mathbf{x}}(i)) = N^{-1}\mathbf{v}_i^T\mathbf{v}_i \quad \textit{for} \quad i = 1, 2, \ldots, L, \tag{47}$$

$$Q_i(\boldsymbol{\mu}(i)) = (N-1)^{-1}\boldsymbol{\mu}^T(i)\boldsymbol{\mu}(i) \quad \textit{for} \quad i = 1, 2, \ldots, L \tag{48}$$

achieve their minimum at the same place.

Here

$$\mathbf{x} = (x_1, \ldots, x_N)^T \tag{49}$$

is a vector of values of the unobserved output,

$$\hat{\mathbf{x}}(i) = (\hat{x}_1(i), \ldots, \hat{x}_N(i))^T \quad \textit{for} \quad i = 1, 2, \ldots, L \tag{50}$$

is a vector of estimates of the states;

$$\mathbf{v} = (x_1 - \hat{x}_1(i), \ldots, x_N - \hat{x}_N(i))^T \quad \textit{for} \quad i = 1, 2, \ldots, L \tag{51}$$

is a vector of filtering errors;

$$\boldsymbol{\mu}(i) = (\mu_1(i), \ldots, \mu_N(i))^T \quad \textit{for} \quad i = 1, 2, \ldots, L \tag{52}$$

is a vector of values of the reconstructed unknown input;

$$\mu_k(i) = W^{-1}(q^{-1}; \mathbf{a})\hat{x}_k(i) = \hat{x}_k(i) - a_1\hat{x}_{k-1}(i) - \ldots - a_n\hat{x}_{k-n}(i) \quad \textit{for} \quad i = 1, 2, \ldots, L,\ k = 1, 2, \ldots, N \tag{53}$$

is a value of the reconstructed input at a time moment k.

Proof. First let us analyse the function

$$Q_i(\boldsymbol{\mu}, \boldsymbol{\mu}(i)) = (N-1)^{-1}\hat{\mathbf{v}}_i^T\hat{\mathbf{v}}_i \quad \text{for} \quad i = 1, 2, \ldots, L. \tag{54}$$

Here

$$\hat{\mathbf{v}}_i = (\mu_1 - \mu_1(i), \ldots, \mu_N - \mu_N(i))^T \quad \text{for } i = 1, 2, \ldots, L, \tag{55}$$

where $\mu_k \neq \mu_k(i)\ \forall\, k = 1, 2, \ldots, N$.

Then

$$\lim_{N\to\infty} N^{-1}\hat{\mathbf{v}}_i^T\hat{\mathbf{v}}_i = \lim_{N\to\infty} N^{-1}\sum_{k=1}^{N}(\mu_k^2 - 2\mu_k\mu_k(i) + \mu_k^2(i))$$
$$= \sigma_\mu^2 + \sigma_{\mu(i)}^2, \quad i = 1, 2, \ldots, L, \tag{56}$$

while

$$\lim_{N\to\infty} N^{-1}\sum_{k=1}^{N}\mu_k\mu_k(i) = \mathrm{cov}(\mu_k\mu_k(i)) = 0, \quad i = 1, 2, \ldots, L, \tag{57}$$

as $\boldsymbol{\mu}_k \sim \mathcal{N}(0, \sigma_\mu^2)$ and $\boldsymbol{\mu}_k(i) \sim \mathcal{N}(0, \sigma_{\mu(i)}^2)\ \forall i = 1, 2, \ldots, L$ are mutually uncorrelated for all k.

Here $\mathrm{cov}(\mu_k\mu_k(i))$ is the covariance between μ_k and $\mu_k(i)\ \forall i = 1, 2, \ldots, L$ and $k = 1, 2, \ldots, N$.

It follows from (56) that function (54) achieves its minimum at the place i and that

$$Q_i(\boldsymbol{\mu}, \boldsymbol{\mu}(i)) < Q_i(\boldsymbol{\mu}, \boldsymbol{\mu}(j)) \quad \text{for} \quad i \neq j, \tag{58}$$

if

$$\sigma_{\mu(i)}^2 < \sigma_{\mu(j)}^2 \quad \text{for} \quad j = 1, 2, \ldots, L-1. \tag{59}$$

Then the functions $Q_i(\boldsymbol{\mu}, \boldsymbol{\mu}(i))$ and $Q_i(\mathbf{x}, \hat{\mathbf{x}}(i))$ for $i = 1, 2, \ldots, L$ acquire their minimum at the same place, as

$$x_k = W(q^{-1}; \mathbf{a})\mu_k \tag{60}$$

and

$$\hat{x}_k(i) = W(q^{-1}; \mathbf{a})\mu_k(i) \quad \text{for} \quad i = 1, 2, \ldots, L,\ k = 1, 2, \ldots, N. \tag{61}$$

It follows from (59) that both functions $Q_i(\boldsymbol{\mu}(i))$ and $Q_i(\mathbf{x}, \hat{\mathbf{x}}(i))$ for $i = 1, 2, \ldots, L$, also have their minimum at the same place.

REMARK 4. The minimal values of above mentioned functions (47) and (48) are unequal.

Conclusion. The relation between the filtering error and the variance of reconstructed input allowed us to replace function (47) of the unknown filtering error by function (48) of the variance $\sigma_{\mu(i)}^2$, $i = 1, 2, \ldots, L$ of reconstructed input $\boldsymbol{\mu}(i)$, $i = 1, 2, \ldots, L$. Then the vector $\hat{\mathbf{x}}(l) = (\hat{x}_1(l), \hat{x}_2(l), \ldots, \hat{x}_N(l))^T$ of optimal estimates of the states

$x_1, x_2, \ldots, x_N$ may be determined using the criterion $Q_i(\boldsymbol{\mu}(i))$ for $i = 1, 2, \ldots, L$ and the condition

$$\hat{\mathbf{x}}(l): \quad Q(\boldsymbol{\mu}(l)) = \min_{\mu(i) \in \Xi} Q_i(\boldsymbol{\mu}(i)) \quad \text{for} \quad i = 1, 2, \ldots, L. \tag{62}$$

Here Ξ is a restricted area of values of the variable $\boldsymbol{\mu}(i)\ \forall\, i = 1, 2, \ldots, L$.

5. State Estimation by Means of the Bank of Parallel Kalman Filters with Optimization

The current state estimates of a linear dynamic system (1), (2) or (11), (12) are calculated using the technique consisting of a bank of the L parallel Kalman filters of the form (24).

Then in order to apply condition (62) in the determination of the vector $\hat{\mathbf{x}}(f) = (\hat{x}_1(f), \hat{x}_2(f), \ldots, \hat{x}_N(f))^T$ of optimal estimates of the states $x_1, x_2, \ldots, x_N$, it is necessary to obtain $\mu_k(i)\ \forall i = 1, 2, \ldots, L$ and $k = n+1, n+2, \ldots, N$ using the equation of the form

$$\mathbf{m}_k = \hat{\mathbf{x}}_k - \mathbf{w}_k \mathbf{a} \quad \forall k = n+1, n+2 \ldots, N. \tag{63}$$

Here

$$\mathbf{m}_k = (\mu_k(1), \ldots, \mu_k(L))^T, \tag{64}$$

$$\hat{\mathbf{x}}_k = (\hat{x}_k(1), \ldots, \hat{x}_k(L))^T, \tag{65}$$

$$\mathbf{w}_k = \begin{bmatrix} \hat{x}_{k-1}(1) & \hat{x}_{k-2}(1) & \ldots & \hat{x}_{k-n}(1) \\ \hat{x}_{k-1}(2) & \hat{x}_{k-2}(2) & \ldots & \hat{x}_{k-n}(2) \\ \ldots & \ldots & \ldots & \ldots \\ \hat{x}_{k-1}(L) & \hat{x}_{k-2}(L) & \ldots & \hat{(x)}_{k-n}(L) \end{bmatrix}, \tag{66}$$

$\mathbf{a}^T = (a_1, \ldots, a_n)$ is a vector of known parameters of polynomial (5); $\mu_k(i)\ \forall i = 1, 2, \ldots, L$ are L estimates of the input value μ_k at a time moment k; $u_k, \ldots, u_{k-n}$ are the values of observed output $\mathbf{U}_k$.

Hence, at the time moment $N+1$ we have a matrix

$$\mathbf{M}_k = \begin{bmatrix} \mu_N(1) & \mu_{N-1}(1) & \ldots & \mu_1(1) \\ \mu_N(2) & \mu_{N-1}(2) & \ldots & \mu_1(2) \\ \ldots & \ldots & \ldots & \ldots \\ \mu_N(L) & \mu_{N-1}(L) & \ldots & \mu_1(L) \end{bmatrix}, \tag{67}$$

the components of which are L estimates of input values $\mu_k\ \forall k = 1, 2, \ldots, N$.

Then, the L estimates of the variance σ_μ^2 are obtained by the formula of the form

$$\begin{bmatrix} \sigma^2_{\mu(1)} \\ \sigma^2_{\mu(2)} \\ \vdots \\ \sigma^2_{\mu(L-1)} \\ \sigma^2_{\mu(L)} \end{bmatrix} = \frac{1}{N-1} \begin{bmatrix} \sum_{i=1}^{N} (\mu_i(1) - \bar{\mu}(1))^2 \\ \sum_{i=1}^{N} (\mu_i(2) - \bar{\mu}(2))^2 \\ \vdots \\ \sum_{i=1}^{N} (\mu_i(L-1) - \bar{\mu}(L-1))^2 \\ \sum_{i=1}^{N} (\mu_i(L) - \bar{\mu}(L))^2 \end{bmatrix}, \tag{68}$$

using, correspondingly, the L reconstructed inputs $\mu_k(i)$ $\forall i = 1, 2, \ldots, L$ and $k = 1, 2, \ldots, N$.

Here $\bar{\mu}(1), \ldots, \bar{\mu}(L)$ are the means of reconstructed inputs, respectively.

If the variance $\sigma^2_{\mu(f)}$ is minimal, i.e.,

$$\sigma^2_{\mu(f)} < \sigma^2_{\mu(i)} \quad \forall i = 1, 2, \ldots, L-1, \tag{69}$$

then the vector $\hat{\mathbf{x}}(f) = (\hat{x}_1(f), \hat{x}_2(f), \ldots, \hat{x}_N(f))^T$ of estimates of the states $x_1, x_2, \ldots, x_N$ is an optimal one according to condition (62). It is obvious that condition (62) and equations (63), (68) let us choose the optimal state estimates $\hat{x}_1(f), \hat{x}_2(f), \ldots, \hat{x}_N(f)$ after processing all observations $u_1, \ldots, u_N$ only. It means that $\hat{x}_k(f)$ for any $k \neq N$ could not be optimal and that at those time moments there exist other optimal $\hat{x}_k(\cdot)$. In order to choose at each time moment k the current optimal state estimate, which would guarantee the current minimal value of reconstructed input and current minimal square filtering error, respectively, at these time moments, it is necessary to rewrite formula (68) in the following form

$$\begin{bmatrix} \sigma^2_{\mu(1)} \\ \sigma^2_{\mu(2)} \\ \vdots \\ \sigma^2_{\mu(L-1)} \\ \sigma^2_{\mu(L)} \end{bmatrix}_k = \frac{1}{k-1} \begin{bmatrix} \sum_{i=1}^{k} (\mu_i(1) - \bar{\mu}(1))^2 \\ \sum_{i=1}^{k} (\mu_i(2) - \bar{\mu}(2))^2 \\ \vdots \\ \sum_{i=1}^{k} (\mu_i(L-1) - \bar{\mu}(L-1))^2 \\ \sum_{i=1}^{k} (\mu_i(L) - \bar{\mu}(L))^2 \end{bmatrix}, \tag{70}$$

and to verify condition (69) for each current time moment k. The recursive formula

$$\begin{bmatrix} \sigma^2_{\mu(1)} \\ \sigma^2_{\mu(2)} \\ \vdots \\ \sigma^2_{\mu(L-1)} \\ \sigma^2_{\mu(L)} \end{bmatrix}_{k+1} = \begin{bmatrix} \sigma^2_{\mu(1)} \\ \sigma^2_{\mu(2)} \\ \vdots \\ \sigma^2_{\mu(L-1)} \\ \sigma^2_{\mu(L)} \end{bmatrix}_{k} + \begin{bmatrix} \varphi_{\mu(1)} \\ \varphi_{\mu(2)} \\ \vdots \\ \varphi_{\mu(L-1)} \\ \varphi_{\mu(L)} \end{bmatrix}_{k+1}$$

$$= \left(1 - \frac{1}{k}\right) \begin{bmatrix} \sigma^2_{\mu(1)} \\ \sigma^2_{\mu(2)} \\ \vdots \\ \sigma^2_{\mu(L-1)} \\ \sigma^2_{\mu(L)} \end{bmatrix}_{k} + \frac{1}{k} \begin{bmatrix} \mu^2_{k+1}(1) \\ \mu^2_{k+1}(2) \\ \vdots \\ \mu^2_{k+1}(L-1) \\ \mu^2_{k+1}(L) \end{bmatrix}. \tag{71}$$

can be used, too. Thus, at each time moment, the minimal variance from (70) can be chosen as well as the state estimate from bank (24) respectivelly. It can be mentioned that in such a case there could appear a false optimum if too small thresholds $\Delta_i \ \forall i = 1, 2, \ldots, L$ would be used in (27).

6. Recursive State Estimation of a Chemical Process

The noiseless sequence x_k is the time-series D from (Box and Jenkins, 1970), which is described by AR(1) model of the form

$$x_k^{(D)} = 1.17 + 0.87 x_{k-1}^{(D)} + \mu_k \quad k = \overline{1, 100}, \tag{72}$$

or by the model of the form

$$\widetilde{x}_k^{(D)} = 0.87 \widetilde{x}_{k-1}^{(D)} + \mu_k \quad k = \overline{1, 100}, \tag{73}$$

if the mean of sequence x_k is eliminated.

Here $x_k^{(D)}, \widetilde{x}_k(D)$ are the values of the above mentioned sequence at a time moment k.

Then the output $\mathbf{U}_k$ to be observed in the presence of outliers (Fig.1) is

$$\widetilde{u}_k^{(D)} = \widetilde{x}_{k-1}^{(D)} + z_k \quad k = \overline{1, 100}, \tag{74}$$

where $\widetilde{u}_k^{(D)}, z_k$ are the values of output $\mathbf{U}_k$ and noise $\mathbf{Z}_k$, respectively, at a time moment k; $\mathbf{Z}_k$ is a sequence of independent identically distributed variables with an 'ε – contaminated' distribution of the form (8) with variance (9).

Both the sequences $\widetilde{\mathbf{x}}_k^{(D)}, \widetilde{u}_k^{(D)}$ from (74) are used for state estimation of the process (73). In this case, for an additive noise $\mathbf{Z}_k$

$$z_k = \begin{cases} 0 & \text{if} \quad \zeta_k > \varepsilon_k, \\ \nu_k 10 & \text{if} \quad \zeta_k < \varepsilon_k, \end{cases} \tag{75}$$

where ν_k, ζ_k are independent Gaussian variables with zero means and variances $1; \varepsilon_k$ is a time varying 'contamination' fraction of the form

$$\varepsilon_k = \begin{cases} 0.1 & \text{for} \quad k = 1, 2, \ldots, 25, \\ 0.2 & \text{for} \quad k = 26, 27, \ldots, 69, \\ 0.05 & \text{for} \quad k = 70, 71, \ldots, 100. \end{cases} \tag{76}$$

For the state estimation by processing $\widetilde{\mathbf{X}}_k^{(D)} = (\widetilde{x}_1^{(D)}, \ldots, \widetilde{x}_{100}^{(D)})^T$ and $\widetilde{\mathbf{U}}_k^{(D)} = (\widetilde{u}_1^{(D)}, \ldots, \widetilde{u}_{100}^{(D)})^T$ the bank of the parallel Kalman filter (24) is used, which can be rewritten for the AR(1) process (73) with the a priori known parameter $a_1 = 0.87$ as

$$\hat{\beta}_i(k+1) = a_1 \hat{\beta}_i(k) + k_{k+1} \psi_i(\widetilde{u}_{k+1}^{(D)} - a_1 \hat{\beta}_i(k)) \quad \text{for} \quad i = \overline{1, 15}, \tag{77}$$

where

$$k_{k+1} = \phi_{k+1}(\hat{\sigma}_z^2 + \phi_{k+1})^{-1}, \tag{78}$$

$$\phi_{k+1} = a_1^2 p_k + \hat{\sigma}_\mu^2, \tag{79}$$

$$p_{k+1} = \phi_{k+1}(1 - k_{k+1}), \tag{80}$$

$$\hat{\beta}_i(k) = \hat{x}_k(i) \quad \text{for} \quad i = \overline{1, 15}, \tag{81}$$

$$\psi_i(e_{k+1}) = \begin{cases} 0 & \text{if} \quad e_{k+1}(i) < -\Delta_i, \\ e_{k+1}(i) & \text{if} \quad -\Delta_i \leqslant e_{k+1}(i) \leqslant \Delta_i \\ 0 & \text{if} \quad e_{k+1}(i) > \Delta_i, \end{cases} \quad \text{for} \quad i = \overline{1, 15}, \tag{82}$$

Moreover, $\Delta_1 = 0.5,\ \Delta_2 = 0.6,\ \Delta_3 = 0.7,\ \Delta_4 = 0.8,\ \Delta_5 = 0.9,\ \Delta_6 = 1,\ \Delta_7 = 1.25,\ \Delta_8 = 1.5,\ \Delta_9 = 2,\ \Delta_{10} = 2.5,\ \Delta_{11} = 3,\ \Delta_{12} = 3.5,\ \Delta_{13} = 4,\ \Delta_{14} = 5,\ \Delta_{15} = 2000$, and $p_0 = 0.1;\ \hat{\sigma}_\mu^2 = 1,\ \hat{\sigma}_z^2 = 1;\ \hat{x}_0 = 0\ \forall I = \overline{1, 15},\ L = 15$.

Then equations (63), (70) unequality (69) were used to obtain optimal state estimates from bank (77) at each time moment k=1,2,...,100.

In Fig. 2, the noiseless and really unobserved output (73) and its three estimates are presented. The first estimate (dotted line) is obtained using only one Kalman filter with $\Delta = 0,5$ in (82), while the second one (dashed line) using the bank of Kalman filters (77) by optimiziting the state estimation itself. The third estimate (dotted-dashed line) is calculated in the absence of the additive noise. For such a case, $\widetilde{u}_k^{(D)}\ \forall k = 1, \ldots, 100$ in (77) is replaced by $\widetilde{x}_k^{(D)}$. In Fig. 3 the operation of the Kalman filters in time for the second and third estimates, respectively, is presented. The dashed line here corresponds to the third estimate, i.e., to the noiseless experiment. From the simulation and state estimation results, presented in Fig. 2, it follows that the accuracy of state estimates, obtained using the bank of the Kalman filters by optimizing the estimation itself, is higher as compared to the accuracy of such estimates based on constant $\Delta = 0.5$, but it is lower, as compared with an accuracy of the estimates, obtained in the absence of additive noise. The results presented in Fig. 3 show that, in such a case, the adaptive technique chooses only one Kalman filter with a threshold $\Delta = 0.9$ in (82) after 2 steps. On the other hand,

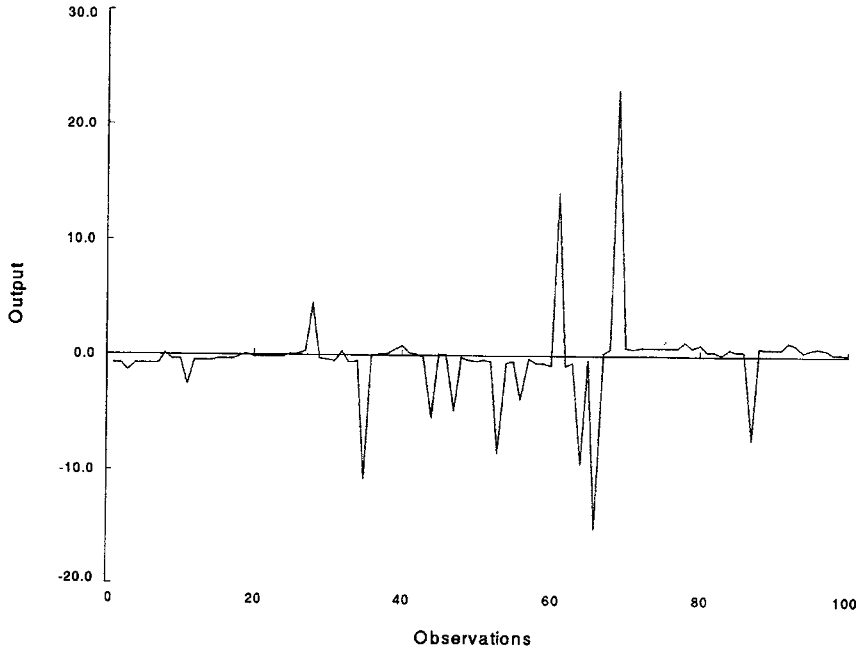

Fig. 1. Time-series D in the presence of time-varying outliers.

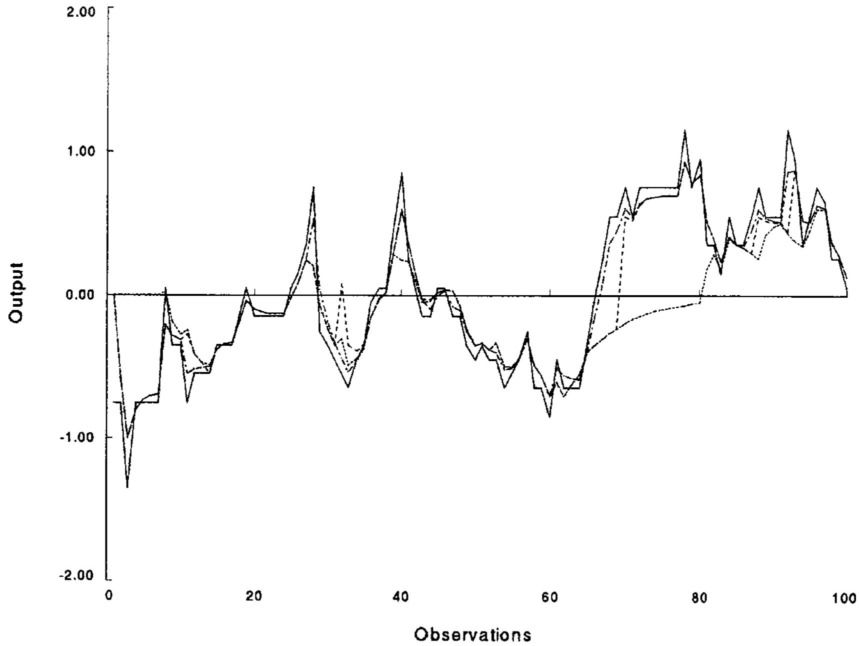

Fig. 2. Noiseless time-series D and its estimates. Estimates: obtained by Kalman filter are denoted as dotted curve; obtained by the bank of filters are dashed and dotted-dashed curves with and without additive noise, respectively.

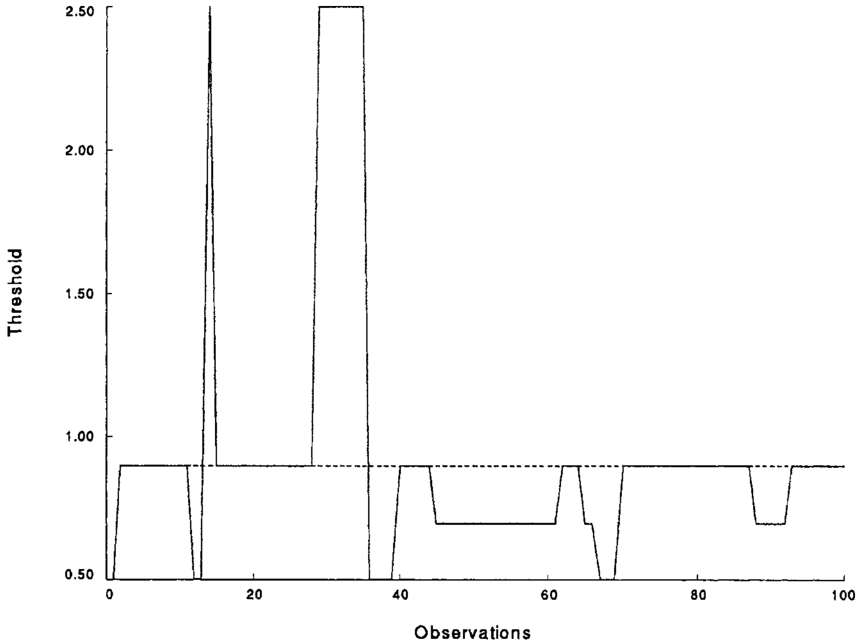

Fig. 3. Operation of the Kalman filters by processing observations. The curves with and without dashes obtained by one Kalman filter and by the bank of the filters respectively.

in the presence of outliers in observations such a technique chooses four Kalman filters at different time moments.

It should be noted that the results averaged by 20 experiments in the presence of time varying outliers and for the case of an unknown parameter $a_1 = 0.87$ in (73) are given in (Pupeikis and Huber, 1997). The banks of the Kalman filters have been used in fault detection, as mentioned in (Zhang, 1989) since 1968 (Newbold and Ho, 1968). Various recursive robust techniques for on-line estimation of dynamic systems parameters in the presence of outliers in additive correlated noise are obtained in Pupeikis, (1994).

7. Conclusions

The classical robust Huber theory of estimation of a location parameter uses stochastic models with time-homogeneous contamination of outliers. In such a case the multivariate recursive robust approaches and techniques are worked out. However, if various robust recursive algorithms turn out to be efficient in the presence of rare and isolated outliers, then there always arise special problems in the presence of time-varying outliers, which often occur in batches. That is why the robust recursive procedures applied in the on-line estimation of states of dynamic processes appeared to be inefficient. In such a case, it is important to solve the generalized problem of a model of outliers, which are varying in time. Therefore this work extends and measurably develops Huber's robust location parameter estimation ideas for linear dynamic processes, which are usually observed in the presence of additive noises, containing time-varying outliers. It is obvious that the new model of outliers requires a new and efficient approach, which could be applied in processing observations in order to obtain the state estimates of the initial dynamic process. In our work such an approach optimiziting the estimation itself has been worked out. Theoretically it is based on a simple but important relation between the variance of reconstructed input of process and one of the filtering errors as well as on the fact that both variables achieve their minimum at the same place. Therefore it is possible to replace the quality-function of unknown filtering error by one of the reconstructed inputs. In practice, our approach is realized by means of the bank of the parallel Kalman filters (24), consisting of simple recursive equations, which differ one from another by threshold Δ_i in the Huber ψ– function (27) only. At each recursive step the current state estimate, which guarantees the minimal filtering error is chosen from a respective bank of current state estimates by the optimization technique, using equations (70) or (71) and condition (69). The results of numerical simulation (Fig. 1–3) using the actual chemical time-series, described by (73), (74), prove the efficiency of the model of time-varying outliers and usefulness of the proposed approach for the state estimation.

Acknowledgment

The author is grateful to Professor Peter J. Huber for helpful comments.

References

Box, G.E.P., and G.M. Jenkins (1970). *Time Series Analysis. Forecasting and Control*. Holden-Day.

Drewelow, W. (1990). *Parameterschätzung nach der Ausgangsfehlermethode*. Messen, Steuern, Regeln. **33**(1), 15–22.

Eykhoff, P. (1974). *System Identification*. Wiley, New York.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986). *Robust Statistics. The Approach Based on Influence Functions*. Wiley, New York.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73–101.

Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.

Ljung, L. (1987). *System Identification. Theory for the User.* Prentice-Hall, Inc.

Masreliez, C.J., and R.D. Martin (1977). Robust bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Trans. on Automatic Control*, **22**(3), 361–371.

Meyr, H., and G. Spies (1984). The structure and performance of estimators for real-time estimation of randomly varying time delay. *IEEE Trans. Acoust.Speech and Signal Process*, **32**, 81–94.

Newbold, P.M., and Y.C. Ho (1968). Detection of changes in the characteristics of a Gauss-Markov process. *IEEE Trans. on Aerospace and Electr. Syst.* **4**(5).

Novovičova, J.(1987). Recursive computation of M-estimates for the parameters of the linear dynamical system. *Problems of Control and Information Theory*, **16**(1), 49–59.

Pupeikis, R., and P.J. Huber (1997). Adaptiv-robuste Zustandsschätzung mit Optimierung des Schätzprozesses. *Bericht zur DFG Projekt ''Echtzeit-Optimierung größer Systeme''*, 82s.

Pupeikis, R. (1994). On-line estimation of dynamic systems parameters in the presence of outliers in observations. *Informatica*, **5**(1–2), 189–210.

Schick, I.C, and S.K. Mitter (1994). Robust recursive estimation in the presence of heavy-tailed observation noise. *Ann. Math. Statist.*, **22**(2), 1045–1080.

Stockinger, N., and R. Dutter (1987). Robust time-series analysis. *Kybernetika*, **23**(1–5).
Verboon, P. (1994). *A Robust Approach to Nonlinear Multivariate Analysis*. Leiden University.
Zhang, X.J. (1989). Auxiliary signal design in fault detection and diagnosis. In M. Thoma and A. Wyner (Eds), *Lecture Notes in Control and Information Sciences*, Vol. 134. Springer.

R. Pupeikis received Ph.D. degree from the Kaunas Polytechnic Institute, Kaunas, Lithuania, 1979. He is a senior researcher at the Process Recognition Department of the Institute of Mathematics and Informatics. His research interest include the classical and robust approaches of dynamic system identification as well technological process control.

DINAMINIŲ SISTEMŲ BŪVIO ĮVERTINIMAS, ESANT STEBĖJIMUOSE NESTACIONARIAM DIDELIŲ IMPULSŲ SRAUTUI

Rimantas PUPEIKIS

Straipsnyje yra nagrinėjami rekurentiniai dinaminių sistemų (1), (2) būvio įvertinimo Kalmano tipo algoritmų (18)–(21) bankai, generuojantys robastinius įverčius, apdorodami stebėjimus, kuriuose yra dideli impulsai, o jų srauto intensyvumas kinta laikui bėgant. Teoriškai (lemos 1, 2, teiginys ir teorema) yra parodyta, kad būvio įverčių vektorius, užtikrinantis minimalią atstatymo ėjimo dispersiją, tuo pačiu garantuoja minimalią filtravimo paklaidą, kuri iš tikrųjų yra nežinoma. Pasiūlytas naujas metodas sistemos būviui įvertinti, grindžiamas robastine filtracija ir optimizavimo procedūra, kurios pagalba kiekvienu laiko momentu iš būvio įverčių banko yra išrenkamas įvertis, optimalus minimalios filtravimo paklaidos prasme. Modeliavimo rezultatai (Pav. 1–3), gauti apdorojus realaus cheminio proceso seką (73), (74), patvirtina teorinių prielaidų pagrįstumą.