LMD mokslo darbai, 447–452
1998, Vilnius

Robust estimation by means of a bank of the Kalman filters

R. Pupeikis (MII)

The Kalman optimal recursive filter applied in the state estimation appeared to be inefficient in the presence of outliers in observations (Masreliez and Martin, 1977). Therefore multivariate robust techniques are worked out (Schick and Mitter, 1994). 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). However, referring to the dynamic discrete-time processes, the disposition of outliers turns out to be very important. In this case it is possible to solve the generalized problem of a model of outliers by means of a bank of the parallel Kalman filters and the procedure of optimization of the state estimation itself, choosing the optimal state estimates with minimal filtering errors (Pupeikis and Huber, 1997).

Assume that we consider the linear discrete-time system of order n with single input and single output described by the autoregressive moving average (ARMA) model of the form

$$x_k = W(q^{-1};\alpha)\mu_k \qquad \forall\ k = 0,1,..., \tag{1}$$

where x_k and μ_k denote the unobserved values of sequences $\{x_k\}$ and $\{\mu_k\}$, respectively, $\{\mu_k\} \sim \mathcal{N}(0,\sigma_\mu^2)$,

$$W(q^{-1};a) = \frac{1-B(q^{-1};b)}{1-A(q^{-1};a)} \tag{2}$$

is a system transfer function,

$$A(q^{-1};a) = \sum_{i=1}^{n} a_i q^{-i}, \tag{3}$$

$$B(q^{-1};b) = \sum_{i=1}^{n} b_i q^{-i} \tag{4}$$

are polynomials,

$$\alpha^T = (a^T, b^T), \qquad a^T = (a_1,...,a_n), \qquad b^T = (b_1,\dots,b_n) \tag{5}$$

are parameters of polynomials (3),(4), $q^{-1}$is the backward shift operator defined by $x_{k-n} = x_k q^{-n}$. Suppose that $\{x_k\}$ is observed under additive noise Z_k, i.e.,

$$u_k = x_k + z_k, \tag{6}$$

where u_k is observed value of output $\mathbf{U}_k$, z_k denotes the unobserved value of the sequence $\{z_k\}$, which is the sequence of independent identically distributed variables with an 'ε-contaminated 'distribution of the form

$$p(z_k) = (1-\varepsilon_k)\ \mathcal{N}(0,\sigma_\xi^2) + \varepsilon_k\ \mathcal{N}(0,\sigma_v^2) \tag{7}$$

and the variance

$$\sigma_z^2 = (1-\varepsilon_k)\sigma_\xi^2 + \varepsilon_k \sigma_v^2, \tag{8}$$

$p(z_k)$ is a probability density distribution of an noise $\mathbf{Z}_k$, moreover, $\sigma_\xi < \sigma_v$ $0 \le \varepsilon_k \le 1$ is the unknown fraction of 'contamination'varying in a time.

It is supposed that the roots of $A(q^{-1};a)$ and $B(q^{-1};b)$ are outside the unit circle of the q^{-1} plane. The true orders n of polynomials (3), (4) and the true values of parameters (5) are known.

The aim of the given paper is a determination of the criterion for optimizing of the robust recursive estimation of states $x_1, x_2, \ldots, x_N$ of the ARMA process (1)–(8) in the presence of time varying outliers in observations $u_1, u_2, \ldots, u_N$ of an output process $\mathbf{U}_k$.

The ARMA model can be rewritten as the stochastic state-space model in the form

$$X_{k+1} = AX_k + h\mu_{k+1}, \tag{9}$$

$$u_k = c^T X_k + d^T w_k + z_k, \tag{10}$$

where

$$X_k = (x_k, x_{k-1}, \ldots, x_{k-n+1})^T \tag{11}$$

is the $n \times 1$ state vector,

$$\mathrm{A} = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \\ 1 & 0 & \ldots & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & \ldots & 1 & 0 \end{bmatrix} \tag{12}$$

is the $n \times n$ matrix, $\mu_{k+1}$is the value of the sequence $\{\mu_k\}$ at time moment $k+1$,

$$h = c = (1,0,\ldots,0)^T, \tag{13}$$

$$d = (b_1, b_2, \ldots, b_n)^T, \tag{14}$$

$$w_k = (\mu_{k-1}, \mu_{k-2}, \ldots, \mu_{k-n})^T \tag{15}$$

are the $n \times 1$ vectors.

So, we have the following structure:

$$\begin{bmatrix} x_{k+1} \\ x_k \\ \cdot \\ \cdot \\ \cdot \\ x_{k-n+2} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \\ 1 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \ldots & 1 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \\ \cdot \\ \cdot \\ \cdot \\ x_{k-n+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \mu_{k+1}, \tag{16}$$

$$u_k = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \\ \cdot \\ \cdot \\ \cdot \\ x_{k-n+1} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & \ldots b_n \end{bmatrix} \begin{bmatrix} \mu_{k-1} \\ \mu_{k-2} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{k-n} \end{bmatrix} + z_k, \tag{17}$$

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

$$\hat{\mathrm{x}}^{(i)}(k+1) = A\hat{\mathrm{x}}^{(i)}(k) + \mathrm{k}(k+1)\ \psi_i(e_{k+1}(i)) \quad \textit{for} \quad i = 1,2,\ldots,L,$$
$$k = 0,1,\ldots,N-1, \tag{18}$$

Here

$$\hat{x}^{(i)}(k+1) = (\hat{x}_{k+1}(i), \hat{x}_k(i), \ldots, \hat{x}_{k-n+2}(i))^T \text{ for } i = 1,2,\ldots,L \tag{19}$$

is the state estimate at time moment $k+1$,

$$\hat{x}^{(i)}(k) = (\hat{x}_k(i), \hat{x}_{k-1}(i), \ldots, \hat{x}_{k-n+1}(i))^T \text{ for } i = 1,2,\ldots,L \tag{20}$$

is the state estimate at k, $\mathrm{k}(k+1)$ is a time-varying filter gain, which is obtained using the respective formulas (Li, 1964) and is the same for all the L filter,

$$\psi_i(e_{k+1}(i)) = \begin{cases} -\Delta_i & if \quad e_{k+1}(i) < -\Delta_i \\ e_{k+1}(i) & if \quad -\Delta_i \le e_{k+1}(i) \le \Delta_i \\ \Delta_i & if \quad e_{k+1}(i) > \Delta_i \end{cases} \quad \text{for } i=1,2,\ldots,L \tag{21}$$

is the special function, which is used to transform the residual $e_k(i)$ for $i=1,2,\ldots,L$ defined by

$$e_k(i) = u_k - c^T \hat{x}_k(i) - d^T w_k \quad \text{for } j=1,2,\ldots,L. \tag{22}$$

The filters in the bank (18) are different because of the threshold Δ_i $\forall$ $i=$ 1,2, ... , L in (21) only, which has a different value $\Delta_1 < \Delta_2 <\ldots< \Delta_L$ for each Kalman filter.

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 oft-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 μ_k will be used.

THEOREM: The functions

$$Q_i(X, \hat{X}^{(i)}) = N^{-1} \mathrm{v}_i^T \mathrm{v}_i \qquad \text{for } i=1,2,\ldots,L\ , \tag{23}$$

$$Q_i(\mu(i)) = (N-1)^{-1}\ \mu^T(i)\mu(i) \qquad \text{for } i=1,2,\ldots,L \tag{24}$$

achieve their minimum at the same place.

Here

$$X = (x_1,\ldots,x_N)^T \qquad \text{for } i=1,2,\ldots,L \tag{25}$$

is a vector of values of an unobserved output,

$$\hat{X}^{(i)} = (\hat{x}_1(i),\ldots,\hat{x}_N(i))^T \qquad \text{for } i=1,2,\ldots,L \tag{26}$$

is a vector of values of the state estimates,

$$\mathrm{v}_i = (x_1 - \hat{x}_1(i),\ldots,x_N - \hat{x}_N(i))^T \quad \text{for } i=1,2,\ldots,L \tag{27}$$

is a vector of filtering errors,

$$\mu(i) = (\mu_1(i),\ldots\mu_N(i))^T \qquad \text{for } i=1,2,\ldots,L \tag{28}$$

is a vector of values of a reconstructed unknown input,

$$\mu_k(i) = W^{-1}(q^{-1};\alpha)\hat{x}_k(i) = \hat{x}_k(i) - \sum_{j=1}^{n} a_j \hat{x}_k(i)q^{-j} + \sum_{j=1}^{n} b_j \mu_k(i)q^{-j} \qquad (29)$$

for $i = 1,2,...,L \quad \forall k = 1,2,...,N$, is a value of a reconstructed input at a time moment k.

Proof First let us analyse the function

$$Q_i(\mu, \mu(i)) = (N-1)^{-1} \hat{\nu}_i^T \hat{\nu}_i \qquad \text{for } i = 1,2,...,L. \quad (30)$$

Here

$$\hat{\nu}_i = (\mu_1 - \mu_1(i),...,\mu_N - \mu_N(i))^T, \qquad \text{for } i = 1,2,...,L, \quad (31)$$

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

Then

$$\lim_{N\to\infty} N^{-1} \hat{\nu}_i^T \hat{\nu}_i = \lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} (\mu_k^2 - 2\mu_k\mu_k(i) + \mu_k^2(i)) = \sigma_\mu^2 + \sigma_{\mu(i)}^2 \text{ for } i = 1,2,...,L, \quad (32)$$

while

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} \mu_k\mu_k(i) = \text{cov}(\mu_k\mu_k(i)) = 0 \qquad i = 1,2,\ldots,L, \qquad (33)$$

as $\mu_k \sim N(0,\sigma_\mu^2)$ und $\mu_k(i) \sim N(0,\sigma_{\mu(i)}^2) \quad \forall i = 1,2,...,L$ are mutually uncorrelated for all k.

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

It follows from (32), that function (30) achieves its minimum at the place l and that

$$Q_i(\mu, \mu(l)) < Q_i(\mu, \mu(j)) \text{ for } l \neq j, \qquad (34)$$

if

$$\sigma_{\mu(l)}^2 < \sigma_{\mu(j)}^2 \text{ for } j = 1,2,...,L-1 . \qquad (35)$$

Then the functions $Q_i(\mu, \mu(i))$ and $Q_i(X, \hat{X}^{(i)})$ for $i = 1,2, ..., L$ acquire their minimum at the same place, as

$$x_k = W(q^{-1};\alpha)\mu_k \qquad (36)$$

and

$$\hat{x}_k(i) = W(q^{-1};\alpha)\mu_k(i) \;\; for \;\; i = 1,2,...,L, \;\forall k=1,2,...,N. \qquad (37)$$

It follows from (32), (34),(35) that the both functions $Q_i(\mu(i))$ and $Q_i(X, \hat{X}^{(i)})$ for $i = 1,2,\ldots, L$, also have their minimum at the same place.

Remark 1. The minimal values of functions (23) and (24) are unequal.

Remark 2. The false minimum appeares if

$$\left|\hat{x}_k(i)\right| < |x_k| \quad \text{for } i = 1,2,\ldots,L, \ \forall k=1,2,\ldots,N, \tag{38}$$

where $m < L$.

Conclusion. The relation between the filtering error and the variance of reconstructed input (29) allowed us to replace the function (23) of the unknown filtering error by the the function (24) of the variance $\sigma^2_{\mu(i)}$ $i = 1,2,\ldots,L$ of a reconstructed input $\mu(i)$ $i = 1,2,\ldots,L$. Then the vector $\hat{x}^{(l)} = (\hat{x}_1^{(l)}, \hat{x}_2^{(l)}, \ldots, \hat{x}_N^{(l)})^T$ of optimal estimates of states $x_1, x_2, \ldots, x_N$ may be determined using criterion $Q_i(\mu(i))$ for $i = 1,2, \ldots, L$ and the condition

$$\hat{x}^{(l)} : Q(\mu(l)) = \min_{\mu(i)\in\Xi} Q_i(\mu(i)) \qquad \textit{for } i = 1,2,\ldots,L. \tag{39}$$

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

REFERENCES

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

[2] Masreliez, C.J., and Martin, R.D.(1977). Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter. *IEEE Trans. on Automatic Control*, AC-22, No. 3, 361-371.

[3] Pupeikis, R. and Huber, P.J. (1997). Adaptiv-robuste Zustandsschaetzung mit Optimierung des Schaetzprozesses. Bericht zur Projekt ''Echtzeit-Optimierung groeßer Systeme'', 82s.

[4] Schick, I.C., and Mitter, S.K. (1994). Robust Recursive Estimation in the Presence of Heavy-tailed Observation Noise. *Ann. Math. Statist.*, Vol 22, No. 2, 1045-1080.

[5] Lee, R.C.K. (1964). *Optimal Estimation, Identification and Control*, M.I.T. Press, Cambridge, Massachusetts.

Robastinis įvertinimas, taikant Kalmano filtrų banką

R. Pupeikis

Straipsnyje nagrinėjamas proceso, aprašomo autoregresijos-slenkančio vidurkio (ARMA) modeliu (1)-(8) būsenų įvertinimo, taikant lygiagrečių Kalmano filtrų banką uždavinys. Įrodyta teorema teigia, kad prieš optimizuojant būsenų įvertinimo procesą galima nežinomas filtravimo paklaidos funkcijos (23) reikšmes pakeist atkurtų įėjimų dispersijų funkcijos (24) reikšmėmis. Po to nesunku spręsti ARMA modelio būsenų robastinio įvertinimo optimizavimo uždavinį, taikant sąlygą (39).