Finite memory observer for input-output estimation. Application to data reconciliation and diagnosis

José Ragot, Frédéric Kratz, Didier Maquin

FINITE MEMORY OBSERVER FOR INPUT-OUTPUT ESTIMATION. APPLICATION TO DATA RECONCILIATION AND DIAGNOSIS

José RAGOT, Frédéric KRATZ, Didier MAQUIN

Centre de Recherche en Automatique de Nancy - UPRESA 7039
Institut National Polytechnique de Lorraine
2, Avenue de la Forêt de Haye - F 54 516 Vandoeuvre les Nancy
Tél.: (33) 3 83 59 59 59 Fax: (33) 3 83 59 56 44

Abstract. We are concerned here with dynamical processes whose states and inputs are measured or partially measured. Due to the noise which corrupt the measurements, we want to simultaneously estimate the states and the inputs. For the case where all the measurements are available, the estimation may be considered as a filtering technique; otherwise, the problem is more complicated as we want also to obtain an estimation of the unmeasured variables. In the very particular but difficult situation where not any inputs is measured, the problem may be linked to those involving unknown input observers; in signal processing, when the system is time-invariant, this problem is often referred to as deconvolution. The proposed technique is applied for measurements corrupted by noise which generally gives bad estimations of the unknown inputs. In order to obtain smooth estimations, additional constraints are proposed.

1. INTRODUCTION

For purpose of simulation, control and optimisation of process, the reliability of data is of great significance. Data reconciliation can be defined as the estimation of measured process data variables to reduce measurement error through the use of temporal and functionnal redundancies. Mathematically, the optipal estimates can be defined as the optimal solution to a constrained least-squares or maximum likelihood objective function. The problem of data reconciliation, in respect to a given model, has been the subject of many works; for specific fields in engineering sciences, data reconciliation is also well known as mass balance or energy balance equilibration. With a general point of view, the underlying techniques deal with constrained optimization theory. However, some others tools are needed, specially observability analysis when only some measurement are available and fault detection procedure where the available data may be corrupted by gross errors [Albuquerque, 1996]. Additionally, variance estimation is somewhat difficult to estimate and the issue of the probability distribution of data cannot be addressed satisfactorily from the observations.

In the specific literature, many reconciliation techniques have been proposed in the steady-state situation with different statistical techniques aiming at the identifications of bias and gross errors. The obvious alternative to steady-state models is to rely on reconciliation methods based on dynamic models. The case of linear systems has been extensively analysed, many of the proposed techniques being rooted on Kalman filtering. However, these methods, probably because they used one-step procedure, at best ameliorate the fluctuations of the measured data, but rarely produce a smooth profile of estimation. Some responses to this problem have been presented by using for example successive linearized horizon and/or neural networks. Despite of these efforts, all the proposed methods have the same aforementioned lack of smoothness. And, as it is well known, this fact is more crucious when an estimation of the input of a process is performed through its measured outputs. One exception has to be formulated concerning the paper of Bagajewicz (1997) which uses an integral approach for dynamic data reconciliation ; this technique allows to obtain, for particular classes of systems, a smoothing effect on the estimated data.

Our paper focuses on presenting a method allowing to estimate the input and the output of a process, in such a way that the estimation satisfy the model of the process, and preserving a smooth profile of the estimations. It is shown that an observer for the process may be designed which works as a sliding-window convolution operator. This kind of observer has also been proposed in [Medvedev 1996], in the case where all the measurements of the process inputs were available.

The paper is organized as follows. First, we formulate the problem of functional estimation. Then, the aforementioned problem is solved by means of a finite memory estimator. Furthermore, the particular case of unknown input is considered. Finally, simulation examples illustrating the proposed estimation procedure are given, one of the example being devoted to diagnosis and fault detection.

2. PROBLEM FORMULATION

In the continuous case, the models of the physical process and of its measurement device are taken as follows:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ x_m(t) = H_x x(t) + b_x(t) \\ u_m(t) = H_u u(t) + b_u(t) \end{cases} \quad (1)$$

where x is the state vector, u and x_m and u_m are respectively the input and the output vectors. Thus, we take into account two measurement equations, the first one for the state and the second one for the input. As particular situation, it is possible to consider the following definitions: $H_x = I$ and/or $H_u = I$. In the following of the paper, one may consider that the matrices H_x and H_u allow to select which components of the state or the input are measured. Thus, the problem consists in finding the estimations $\hat{x}(t)$ and $\hat{u}(t)$ of the states and the inputs from the available measurements on the time window $\lfloor t_0,\ t_1 \rfloor$. The system is supposed to be estimable and the corresponding conditions will be given further.

Assuming a normal distribution for the measurement errors with variance matrices V_u and V_x, it is possible to simultaneously estimate the states and the inputs by using the likelihood maximum principle, that consists in minimizing in respect to $\hat{x}$ and $\hat{u}$ the quadratic criterion:

$$\Phi = \frac{1}{2}\int_{t_0}^{t_1} \left(\begin{array}{c} \left\| H_x \hat{x}(t) - x_m(t) \right\|^2_{V_x^{-1}} + \\ \left\| H_u \hat{u}(t) - u_m(t) \right\|^2_{W_u^{-1}} \end{array} \right) dt \quad (2a)$$

under the linear constraints:

$$\dot{\hat{x}}(t) - A\hat{x}(t) - B\hat{u}(t) = 0 \qquad t \in \left[t_0,\ t_1 \right] \quad (2b)$$

The associated cost function may be expressed as:

$$L = \frac{1}{2} \left(\begin{array}{c} \left\| H_x \hat{x}(t) - x_m(t) \right\|^2_{V^{-1}} + \\ \left\| H_u \hat{u}(t) - u_m(t) \right\|^2_{W^{-1}} \end{array} \right) + \lambda^T \left(\dot{\hat{x}}(t) - A\hat{x}(t) - B\hat{u}(t) \right) \quad (3)$$

and the optimality conditions:

$$\begin{cases} H_x^T V^{-1} \left(H_x \hat{x}(t) - x_m(t) \right) - A^T \lambda(t) - \dot{\lambda}(t) = 0 \\ H_u^T W^{-1} \left(H_u \hat{u}(t) - u_m(t) \right) - B^T \lambda(t) = 0 \\ \dot{\hat{x}}(t) = A\hat{x}(t) + B\hat{u}(t) \end{cases} \quad (4)$$

In order to be able to solve these equations, they have to be completed with boundary conditions on the states. Thus we are involved with a classical but difficult problem with two side conditions. Generally, the resolution may be obtained only through a numerical approach due to the fact that the measurements are given with a sampling period.

3. ESTIMATION PROCEDURE

Finding the optimum of the estimation criterion (2a) being more easier with a discrete representation, the formulation of the problem is now adapted to the case of sampled data. For that, the following models are used:

$$\begin{cases} x(k+1) = \Phi x(k) + \Psi u(k) \\ x_m(k) = H_x x(k) + b_x(k) \\ u_m(k) = H_u u(k) + b_u(k) \end{cases} \quad (5)$$

On the observation window $\lfloor 1,\ N \rfloor$, the model (5) may be written more compactly:

$$\begin{cases} A\hat{X} + B\hat{U} = 0 \\ X_m = CX + E_x \\ U_m = DU + E_u \end{cases} \quad (6)$$

where $\hat{X}$ and $\hat{U}$ are formed, by piling according to the time the different states and inputs:

$$X = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{pmatrix} \qquad U = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{N-1} \end{pmatrix}$$

and:

$$A = \begin{pmatrix} -\Phi & I & 0 & 0 & \\ 0 & -\Phi & I & 0 & 0 \\ . & . & & & \\ 0 & 0 & 0 & -\Phi & I \end{pmatrix} \quad C = \begin{pmatrix} H_x & & \\ & \dots & \\ & & H_x \end{pmatrix}$$

$$B = \begin{pmatrix} \Psi & 0 & 0 & 0 \\ 0 & \Psi & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \Psi \end{pmatrix} \qquad D = \begin{pmatrix} H_u & & \\ & \dots & \\ & & H_u \end{pmatrix}$$

Assuming a normal distribution of the measurement errors, the criterion to be minimized in respect to $\hat{X}$ and $\hat{U}$ may be expressed:

$$\Phi\left(\hat{X}, \hat{U} \right) = \left\| C\hat{X} - X_m \right\|^2_{V^{-1}} + \left\| D\hat{U} - U_m \right\|^2_{W^{-1}} \quad (7)$$

where V and W directly depend on V_x and V_u. In order to reduce the influence of noise and to stabilise the solution, a regularized-least-squares method has been used in this study [Tu, 1997]. This is also motivated to the necessity of satisfying a certain level of smoothness for the inputs and states estimations. Therefore, to find a solution which is also consistent with all the measurement data, the following expression must be minimized :

$$\tilde{\Phi} = \left\| C\hat{X} - X_m \right\|^2_{V^{-1}} + \left\| D\hat{U} - U_m \right\|^2_{W^{-1}} + \alpha\varphi\left(\hat{X} \right) + \beta\varphi\left(\hat{U} \right)$$

This form shows that a compromise is considered between the fidelity to the data and the roughness of the estimate. The structure of the φ function has to be

selected by the user. Here are two convenient examples:

$$\varphi\left(\hat{X}\right) = \left\|E\hat{X}\right\|^2 \qquad \varphi\left(\hat{U}\right) = \left\|F\hat{U}\right\|^2 \tag{8a}$$

$$\varphi\left(\hat{X}\right) = \frac{\left\|E\hat{X}\right\|^2}{1+\left\|E\hat{X}\right\|^2} \qquad \varphi\left(\hat{U}\right) = \frac{\left\|F\hat{U}\right\|^2}{1+\left\|F\hat{U}\right\|^2} \tag{8b}$$

The matrices E and F are chosen in order to introduce a filtering effect between two (or more) consecutive data. The values of α and β (the so-called regularization parameters) determine the relative weight between the fidelity to experimental data and the smoothness of the solution. If they are too small, the regularization effect vanishes and an ill-conditioned solution is approached. On the other hand, too large values produce smooth estimates that may be unable to track sharp variations in the true inputs and states (a discussion about the choice of these parameters may be found in Twomey, 1965) and interpretation of the regularization scheme has been examined in (Commenges, 1984). Specific applications in the filed of physiological systems arev reported in [Nicolao, 1997]. When roughness is measured by the sum of squared second differences, one just uses

$$E = \begin{pmatrix} 1 & 0 & & & \\ -2 & 1 & 0 & & \\ 1 & -2 & 1 & & \\ \cdots & & & & \\ 0 & & & 1 & -2 & 1 \end{pmatrix}$$

In some cases, penalizing the second difference leads to solutions that are excessively smooth, and better results are obtained by penalizing the estimated itself or its first difference. In such case one may choose $E = I$ or

$$E = \begin{pmatrix} 1 & 0 & & & \\ -1 & 1 & 0 & & \\ 0 & -1 & 1 & & \\ \cdots & & & & \\ 0 & & & 0 & -1 & 1 \end{pmatrix}$$

The choice of the regularization parameters have been the subject of several studies that have been reported in the literature on spline approximation ; however, many ideas from spline-approximation resultas can be extended, more or less directly, to the input-output problem. The well known technique of Lagrange's multipliers is used in order to obtain the estimation; with the regularisation (8a), it can be easily established the following optimality equations:

$$C^T V^{-1}\left(C\hat{X} - X_m\right) + A^T\lambda + \alpha E^T E\hat{X} = 0 \tag{9a}$$

$$D^T W^{-1}\left(D\hat{U} - U_m\right) + B^T\lambda + \beta F^T F\hat{U} = 0 \tag{9b}$$

$$A\hat{X} + B\hat{U} = 0 \tag{9c}$$

This set of equations being linear in respect to $\{\lambda,\ X,\ U\}$, its numerical resolution is straightforward. In order to expressed an analytical solution, an elimination procedure of the Lagrange parameters may be used. First, let us define the matrices:

$$P = C^T V^{-1} C + \alpha E^T E + A^T A \tag{10a}$$

$$Q = D^T W^{-1} D + \beta F^T F + B^T B \tag{10b}$$

If the matrices $\begin{pmatrix} C \\ E \\ A \end{pmatrix}$ and $\begin{pmatrix} D \\ F \\ B \end{pmatrix}$ are full column rank, then the matrices P and Q are regular. Thus, after some calculations, it is possible to express the estimations as:

$$\begin{aligned} \hat{X} = {} & R^{-1}P^{-1}\left(\left(C^T V^{-1} - A^T\left(I - T_B\right)H^{-1}G_x\right)\right)X_m \\ & - R^{-1}P^{-1}A^T\left(\left(I - T_B\right)H^{-1}G_u + BQ^{-1}D^T W^{-1}\right)U_m \end{aligned} \tag{11a}$$

$$\begin{aligned} \hat{U} = {} & S^{-1}Q^{-1}\left(\left(D^T W^{-1} - B^T\left(I - T_A\right)H^{-1}G_x\right)\right)U_m \\ & - S^{-1}Q^{-1}B^T\left(\left(I - T_A\right)H^{-1}G_u + AP^{-1}C^T V^{-1}\right)X_m \end{aligned} \tag{11b}$$

with the definitions:

$$\begin{cases} T_a = AP^{-1}A^T & T_b = BP^{-1}B^T \\ R = I - P^{-1}A^T T_B A & S = I - Q^{-1}B^T T_A B \\ G_x = \left(AR^{-1} - BS^{-1}Q^{-1}B^T A\right)P^{-1}C^T V^{-1} & \\ G_u = \left(BS^{-1} - AR^{-1}P^{-1}A^T B\right)Q^{-1}D^T W^{-1} & \\ H = AR^{-1}P^{-1}A^T\left(I - T_B\right) + BS^{-1}Q^{-1}B^T\left(I - T_A\right) & \end{cases} \tag{12}$$

Thus, the expressions (11), with the help of definitions (10) and (12), allow to estimate, from the available measurements, the states and the inputs of the considered process. Despite of the apparent complexity of the results, eq. (11) clearly shows that the proposed structure looks like a finite memory observer. The obtained estimations may be used either on a fixed time window, or a sliding time window (in this last case, it should be point out that the gain of the filter between the measurements and the estimated has to be computed in a preliminary step and is constant). Moreover, one can say that the expressions (11) are those of a fixed memory filter whose design parameters, besides the length of the time window, are the filtering coefficients α and β; the last ones are left to the user and may be chosen according to the noise level.

It is clear that the size of the optimization problem increases linearly with the number of measurements. For on line diagnosis purpose, an estimation technique has to computationally feasible and a consequently we must be able to bound the number of variables to be estimated. This motivate to use a fixed-size moving window in which the number of measurements that we base our estimate on (and, hence the size of the optimization) remains constant. Considering eq. (11), it appears that the gains of the observer remain constant and be computed off line. Concerning the size of the window, a compromise has to be performed. The win-

dow has to contain enough information to obtain suitable estimates of the states ; however, for on-line applications, low signal-to-noise ratios, there is a need to reduce the computational requirements (i.e. the horizon size). An additional advantage to the moving window approach is the memory effect : since the window only extends backwards in time for a finite length of time, data collected long before the current time does not enter into the estimation for the current time.

4. THE UNMEASURED INPUTS CASE

It is an extreme situation where the input of a process has to be reconstructed only from its output. However, it is a common situation in the field of automatic control and also in signal processing. For that purpose, the proposed technique may be directly applied when considering the limit case where W tends toward infinity.

$$\hat{X} = R^{-1}P^{-1}\left(\begin{pmatrix} C^T V^{-1} - \\ A^T\left(I - BQ^{-1}B^T\right)H^{-1}G_x \end{pmatrix}\right)X_m$$

$$\hat{U} = -S^{-1}Q^{-1}\left(\begin{pmatrix} B^T AP^{-1}C^T V^{-1} + \\ B^T\left(I - AP^{-1}A^T\right)H^{-1}G_x \end{pmatrix}\right)X_m$$

with the new definition of the Q matrix :

$$Q = \beta F^T F + B^T B$$

5. EXAMPLES

Let us consider the first order system :

$$\begin{cases} x_{k+1} = 0.9x_k + 0.1u_k \\ x_{m,k} = x_k + b_k \end{cases}.$$

where only the state has been measured. The test data obtained from this process are shown in figure 1. The figure shows the following results (from top to bottom):

the measured input (is not used in the estimation procedure)
the measured output (corrupted by noise)
the estimated input and the measured input
the estimated state and the measured state
the estimated input and the true input
the estimated state and the true state

As it can be seen, although the measurement input has not been taken into account, a very good reconstruction of the input has been obtained.

Application to process diagnosis

In order to detect faults affecting a process, the well-known technique of residual structuration may be considered Analogous to the employment of analytical observers in fault detection and isolation [Patton, 1989], the proposed finite memory observer (FMO) is able to serve as a substitute in the case where no complete information is available or used. Finite Memory observer scheme (FMDOS) and Finite Memory Generalized Observer Scheme (FMGOS) may be used to isolate faults. In FMDOS each estimation is driven by an output from a different single sensor, and all the variables (input, state, output) are estimated. In FMGOS, an estimator is dedicated to a certain sensor and is driven by all outputs except that of the respective sensor. The same technique may be defined and applied to actuator fault detection and isolation. The following process has been used to simulate data :

$$x_{k+1} = \begin{pmatrix} -0.25 & .925 \\ -.5 & .95 \end{pmatrix}x_k + \begin{pmatrix} .75 & -.25 \\ 0 & -.5 \end{pmatrix}u_k$$

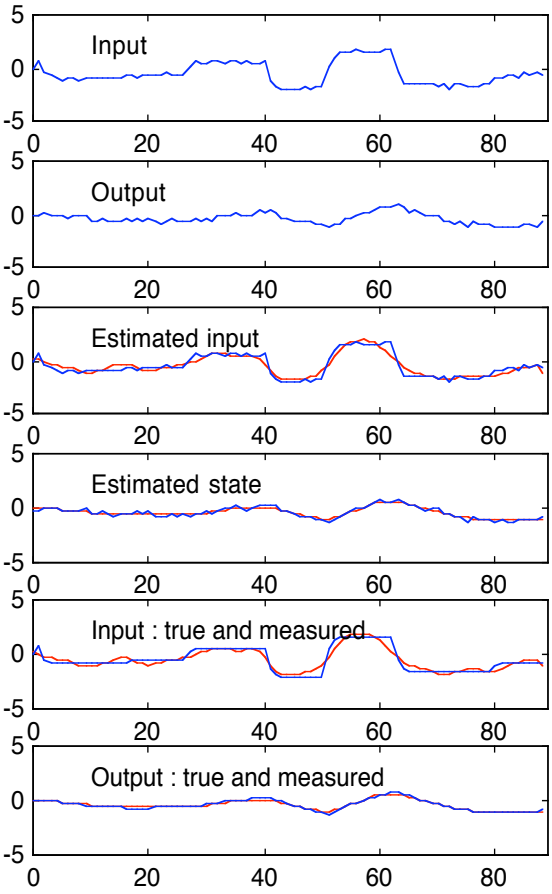

Figure 1. Unknow input estimation

During the simulation, faults have been added. Two constant actuators faults were simulated between sample time 116 and 118 for the first actuator and between samples time 56 and 62 for the second actuator. The sensor faults are also constant and appear between sample time 84 and 90 for the first sensor and between 28 and 34 for the second.

All the inputs and the outputs are measured, but according to the FMDOS and FMGOS schemes, different selections of measurement will be considered. We have examined five estimators depending on this selection :

	Inp. 1	Inp. 2	Out. 1	Out. 2
case 1	x	x	x	x
case 2	x	x	x	0
case 3	x	x	0	x
case 4	x	0	x	x
case 5	0	x	x	x

Table 1.

On the figures 2 to 6, corresponding to the five cases of table 1, the results are organized as follows. The two first parts of each figure show the measured inputs and their estimations allowing to generate two residuals ; the two last parts are related to the outputs (or state variables) and their estimations, allowing to generate two other residuals. The vertical dashed lines indicate the time duration of the faults.

It can be observed the following results. When all the measurements are used (fig. 2), the residuals are affected by the four faults. However, the sensitivity of the residuals to the faults are different (see for example, the residual on input 2 which is very sensitive to fault on actuator 2, while the other residuals have very small variations according to this fault). As a general result, it can be seen and proved that a residual is mostly affected by the fault on its corresponding variables.

When only a part of the measurements is used, the results have the following interpretation. For example, when using all the measurements except the first state (fig. 3), the residual on input 2 is sensitive to the fault on this input ; the other residuals and more or less sensitive to this fault. Same constatation and justification may be given in the other situations (fig. 4, 5 and 6). As it is well known in fault detection and isolation techniques with infinite memory observer, a decision table may be established in order to isolate the faults.

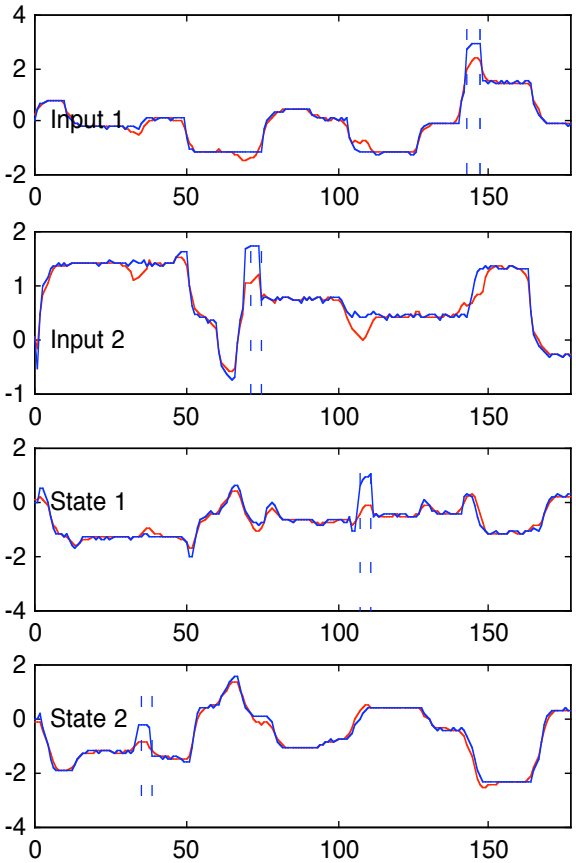

Figure 2. Estimation. Case 1.

CONCLUSION

This paper has presented a global approach to dynamic data reconciliation of linear systems ; by comparison with approaches taking into account only the output measurements, the propose technique performs an estimation of both input, output and state variables. Two applications have been pointed out : the estimation of an unknown input and the detection-localisation of faults. Among the further works is now being undertaken the analysis of the parameters of the observer : the length of the sliding window, the weight allowing to adjust the smoothing effect of the estimations.

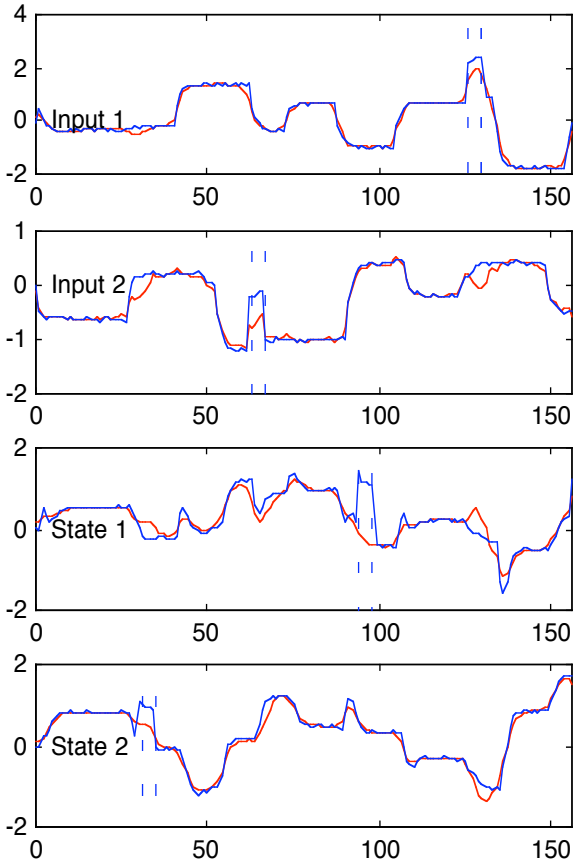

Figure 3. Estimation. Case 2.

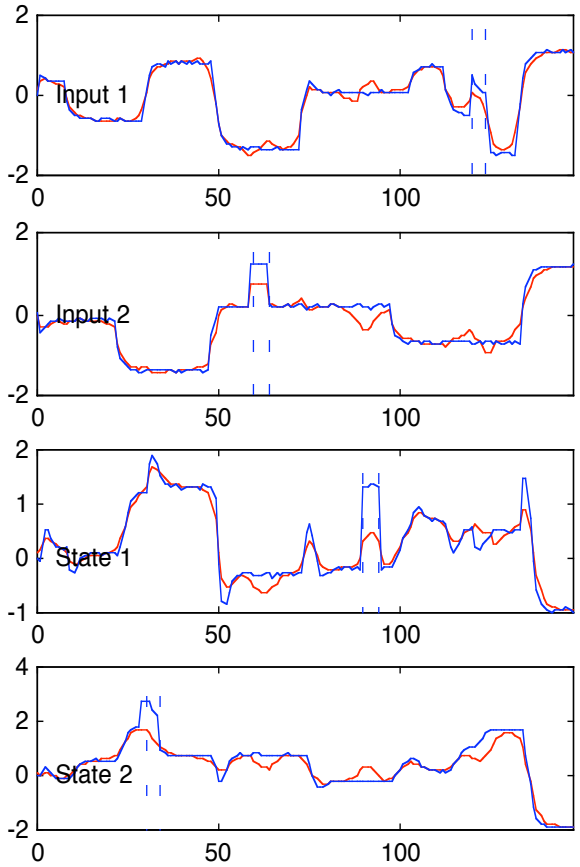

Figure 4. Estimation. Case 3.

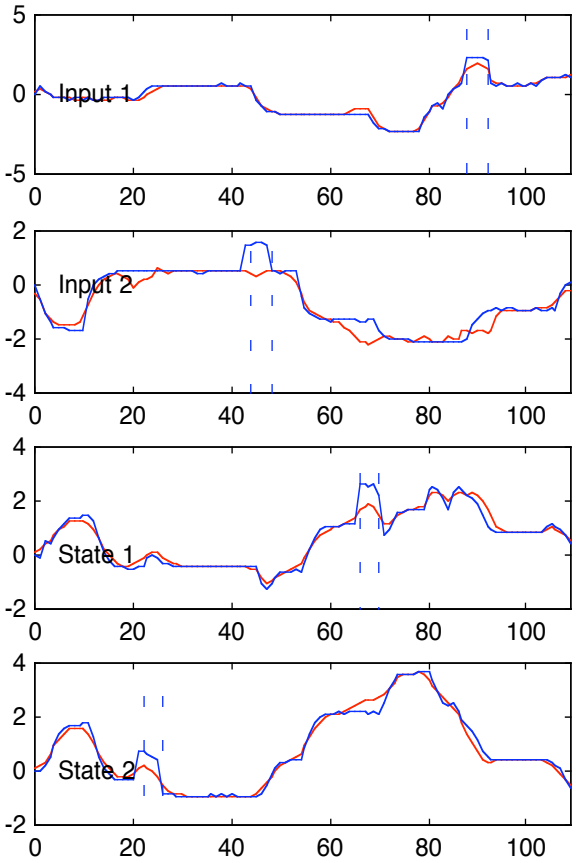

Figure 5. Estimation. Case 4.

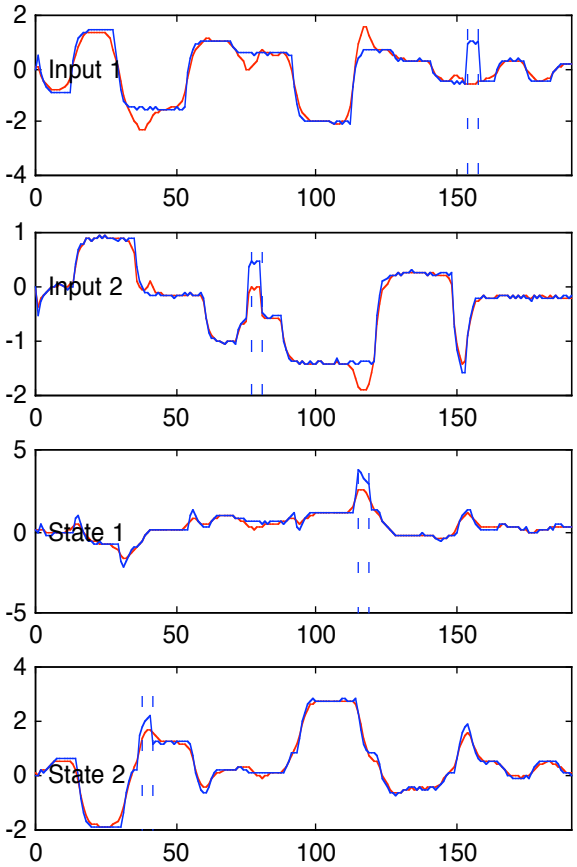

Figure 6. Estimation. Case 5.

REFERENCES

Albuquerque J.S., L.T. Biegler (1996). Data reconciliation and gross-error detection for dynamic systems. AIChE journal, 42, p. 2841.

Bagajewicz M.J., Q. Jiang (1997). Integral approach to plant linear dynamic reconciliation. AIChE Journal, 43 (10), p. 2546-2558.

Corless M., J. Tu (1998). State and output estimation for a class of uncertain system. Automaztica, 34 (8), p. 757-763.

Commenges D (1984). Approche bayésienne en traitement de signal : estimation du signal d'entrée d'un système linéaire. Revue de statistique appliquée, 30, p. 13-22.

Hostetter G., J.S. Meditch (1973). Observing systems with unmeasurable inputs. IEEE Transactions on Automatic Control, AC-18, p. 307-308.

Kratz F., W. Nuninger, J. Ragot (1997). Finite memory observer based method for failure detection in dynamic systems. 11th IFAC Symp. on System Identification, SYSID'97, Fukuoka, Japan.

Liebman M.J., T.F. Edgar, L.S. Lasdon (1992). Efficeint data reconciliation and estimation for dynamic process usnig non linear programming techniques. Computers chemical engineering, 16 (40), p. 963-986.

Nuninger W., F. Kratz, J. Ragot (1998). Finite memory generalized state observer for failure detection in dynamic systems. IEEE Conf. on Decision and Control, CDC'98, Tampa, Florida, USA.

Maquin D.., J. Ragot (1991). Comparison of gross errors detection methods in process data. 30th Conf. on Decision and Control IEEE, Brighton.

Medvedev A (1996). Continuous least-squares observers with applications. IEEE Transactions on Automatic Control, 41 (10), p. 1530-1537.

De Nicolao G., G. Sparacino, C. Cobelli. Nonparametric input estimation in physiological systems: problems, methods and case studies. Automatica, 33 (5), p. 851-870, 1997.

Patton R.J., P.M. Frank, R.N. Clark (1989). Fault diagnosis in dynamic systems. Prentice Hall.

Twomey S (1965). The application of numerical filtering to the solution of integral equations of the first and encountered in indirect sensing measurements. Journal of Franklin Institute, 279, p. 95-109.

Tu Y.X., A. Wernsdörer, S. Honda, Y. Tomita (1999). Estimation of conduction velocity distribution by regularized-least-square method. 11th IFAC Symposium on System Identification, SYSID, p. 1303-1308.