Data reconciliation : a robust approach using contaminated distribution

José RAGOT, Mohammed CHADLI, Didier MAQUIN
E-mail: {jragot, mchadli, dmaquin}@ensem.inpl-nancy.fr
Centre de Recherche en Automatique de Nancy. CNRS UMR 7039
Institut National Polytechnique de Lorraine. 2, Avenue de la forêt de Haye. F - 54 516 Vandoeuvre les Nancy

Abstract - On-line optimisation provides a means for maintaining a process around its optimum operating plant. An important component of optimisation relies in data reconciliation which is used for obtaining consistent data. On a mathematical point of view, the formulation is generally based on the assumption that the measurement errors have normally pdf with zero mean. Unfortunately, in the presence of gross errors, all of the adjustments are greatly affected by such biases and would not be considered as reliable indicators of the state of the process. This paper proposes a data reconciliation strategy that deals with the presence of such gross errors.

1. Introduction

The problem of obtaining reliable estimates of the state of a process is a fundamental objective, these estimates being used to understand the process behaviour. For that purpose, a wide variety of techniques has been developed to perform what is currently known as data reconciliation. Unfortunately, the measurement may be unknowingly corrupted by gross errors. As a result, the data reconciliation procedure can give rise to absurd results and the estimated variables are corrupted by this bias. Several schemes have been suggested to cope with the corruption of normal assumption of the errors [Narasimhan, 1989]. Methods to include bounds in process variables to improve gross errors detection have been developed. One major disadvantage of these methods is that they give rise to situations that it may impossible to estimate all the variable using only a subset of the remaining free gross errors measurements.

There is also an important class of robust estimators whose influence function are bounded and finit allowing to reject outliers [Hampel, 1986]. Another approach is to take into account the non ideality of the measurement error distribution using an objective function constructed on contaminated error distribution.

In the following, we adopt and develop this idea for the data reconciliation problem. Section 2 will be devoted to recall the background of data reconciliation. In section 3, robust data reconciliation is developped and will be illustrated through an academic example in section 4.

2. Data reconciliation background

The classical general data reconciliation problem [Mah, 1976], [Crowe, 1996], deals with a weighted least squares minimisation of the measurement adjustments subject to the model constraints. Indeed the model process equations are taken as linear for sake of simplicity :

$$Ax = 0,\ \ A \in \Re^{m.n},\ \ x \in \Re^n \tag{1}$$

where x is the state of the process. The measurement devices give the information $\tilde{x} \in \Re^n$:

$$\tilde{x} = x + \varepsilon,\ \ p(\varepsilon) \approx \mathrm{N}(0, V) \tag{2}$$

where $\varepsilon \in \Re^n$ a vector of random errors characterised by varaince matrix V and normal probability distribution (pd). In the least square sense, the well-known solution of this problem is $\hat{x} = (I - VA^T(AVA^T)^{-1}A)y$. [Maquin, 1991]. In fact, the method doesn't work in any situation, the main drawback being the contamination of all estimated values by the outliers. For that reason robust estimators could be preferred, robustness being the ability to ignore the contribution of extreme data i.e. such as gross errors.

3. Robust data validation

If the measurements contain random outliers, then a single pd described as in (2) cannot account for the high variance of the outliers. To overcome this problem let us assume that measure noise is sampled from two pd, one having a small variance representing regular noise and the other having a large variance representing outliers. Thus, for each observation, we define:

$$p_{j,i} = \frac{1}{\sqrt{2\pi}\sigma_j}\exp\left(-\frac{1}{2}\left(\frac{x_i - \tilde{x}_i}{\sigma_j}\right)^2\right),\ \ j = 1,2, \quad i = 1,..,n$$

$$p\left(x_i \middle| y_i, \theta\right) = \mu p_{1,i} + (1-\mu)p_{2,i}$$

allowing to define the likelihood function :

$$\Phi = \prod_{i=1}^{n} p\left(x_i \middle| \tilde{x}_i, \theta\right) \tag{3}$$

Minimising (3) in respect to x gives :

$$x = \left(I - W_x A^T \left(AW_x A^T\right)^{-1} A\right)\tilde{x} \tag{4a}$$

$$W_x^{-1} = diag\left(\frac{\frac{\mu}{\sigma_1^2}p_{1,i} + \frac{1-\mu}{\sigma_2^2}p_{2,i}}{\mu p_{1,i} + (1-\mu)p_{2,i}}\right) \tag{4b}$$

Thus system (4) is clearly non linear and we suggest the following iterative scheme :

$$x^{(0)} = \tilde{x} \tag{5a}$$

$$p_j^{(k)} = \frac{1}{\sqrt{2\pi}\sigma_j}\exp\left(-\frac{1}{2}\left(\frac{x^{(k)} - \tilde{x}}{\sigma_j}\right)^2\right),\ \ j = 1,2 \tag{5b}$$

$$x^{(k+1)} = \left(I - W_x^{(k)} A^T \left(AW_x^{(k)} A^T\right)^{-1} A\right)\tilde{x} \tag{5c}$$

4. Example and discussion

The method described in section 3 is applied to system (6). Random errors were added to the 16 variables but the gross errors were added only on some of them.

$$
\begin{aligned}
&x_1 - x_2 + x_4 = 0 && x_2 - x_3 - x_{11} = 0 \\
&x_3 - x_4 - x_5 = 0 && x_5 - x_6 + x_{10} = 0 \\
&x_6 - x_7 - x_8 = 0 && x_7 - x_9 - x_{10} = 0 \\
&x_{12} + x_{13} - x_{14} = 0 && x_{14} - x_{15} - x_{16} = 0 \\
&x_{11} - x_{12} - x_{13} + x_{16} = 0 &&
\end{aligned}
\qquad (6)
$$

The performance results are given when three gross errors (with a common magnitude of 8) affect the measurement 3, 7 and 16. Comparison of the proposed robust least square algorithm (RLS) with the classical least squares (LS) algorithm is now provided.

Var.	true data	meas.	RLS est.	LS est.
1	115.00	114.50	114.26	114.91
2	132.00	129.80	130.55	132.05
3	106.40	114.53	105.26	108.28
4	17.00	17.04	16.28	17.14
5	89.40	88.37	88.97	91.14
6	109.60	110.90	110.05	111.86
7	61.00	69.72	61.74	63.56
8	48.60	48.58	48.30	48.29
9	40.80	40.93	40.66	42.85
10	20.20	20.23	21.08	20.72
11	25.60	25.55	25.29	23.77
12	33.10	33.34	33.54	36.97
13	5.10	5.07	5.27	5.15
14	38.20	39.03	38.82	42.13
15	25.60	25.56	25.29	23.77
16	12.60	20.61	13.53	18.36

Table 1. Measurements and reconciled data

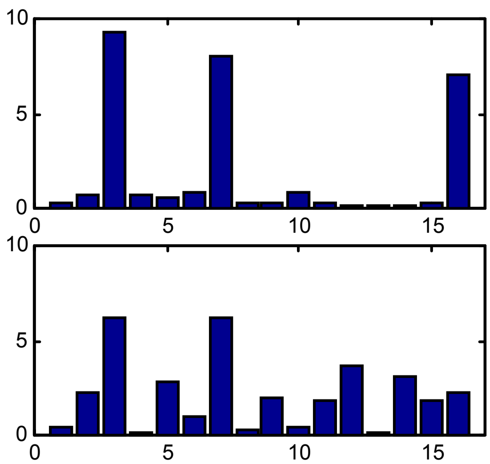

Figure 1. Corrective terms for RLS and LS

In table 1, columns 4 and 5 show the estimations obtained with RLS and LS ; analysing the estimation errors, for RLS estimator clearly allows to suspect variables 3, 7 and 16 for being contaminated by a gross error. Such conclusion is more difficult to express with LS estimator. Figure 1 visualizes more clearly the estimation errors both for LS and RLS (on each graph, horizontral and vertical axis are scaled with the number of the data and the magnitude of the absolute estimation error). The proposed method has been extended to non linear system. We relate only the bilinear case, in which the model is described by : $Ax = 0$, $A(x \otimes y) = 0$ with $x \in \Re^n$, $y \in \Re^n$, $A \in \Re^{m.n}$. The criterion to be maximised is then defined by

$$\Phi = \prod_{i=1}^{v} p\left(x_i \middle| \tilde{x}_i, \theta\right) p\left(y_i \middle| \tilde{y}_i, \theta\right) \qquad (7)$$

Numerical data are not given and only graphical results are shown with figure 2 : the upper part is concerned with estimation errors obtained with RLS while lower part is devoted to LS. For that simulation the x and y data were respectively corrupted with gross errors on component 3, 7, 16 on x and 1, 9, 12 on y. Without ambiguity, all the gross errors have been detected and isolated with RLS that is not the case with LS.

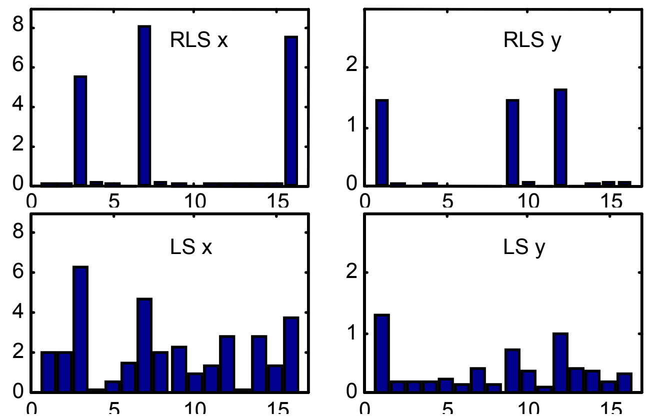

Figure 2. Corrective terms for RLS and LS

5. Conclusion

To deal with the issues of gross errors influence on data estimation, the paper has presented a robust approach. For that purpose, we use a cost function which is less sensitive to the outlying observations than that of least squares. As a perspective of development of robust reconciliation strategies, there is a need for taking account of model uncertainties and optimise the balancing parameter μ .

References

[1] N. Arora. Redescending estimator for data reconciliation and parameter estimation. *Computers and Chemical Engineering*, 25, p. 1585, 2001.

[2] C.M. Crowe. Data reconciliation - progress and challenges. *Journal Process Control*, 6 (2,3), p. 89-98, 1996.

[3] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stohel. Robust statistic : the approach on influence functions. Wiley, New-York, 1986.

[4] R.S.H. Mah, G.M. Stanley and D. Downing. Reconciliation and rectification of process flow and inventory data, *Ind. Eng. Chem*., Process Des. Dev., 15 (1), 1976.

[5] D. Maquin, G. Bloch, J. Ragot. Data reconciliation for measurements. *Europeen journal of diagnosis and safety in automation*, 1 (2), p. 145-181, 1991.

[6] S. Narasimhan, R.S.H. Mah. Treatment of general steady state process models in gross error identification. *Computers and Chemical Engineering*, 13 (7), p. 851-853, 1989.