Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Detection of gross errors using mixed integer optimization approach in process industry*

MEI Cong-li†, SU Hong-ye†, CHU Jian

(*National Laboratory of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou 310027, China*)

†E-mail: clmei@iipc.zju.edu.cn; hysu@iipc.zju.edu.cn

Received Sept. 5, 2006; revision accepted Jan. 8, 2007

Abstract: A novel mixed integer linear programming (NMILP) model for detection of gross errors is presented in this paper. Yamamura *et al.*(1988) designed a model for detection of gross errors and data reconciliation based on Akaike information criterion (AIC). But much computational cost is needed due to its combinational nature. A mixed integer linear programming (MILP) approach was performed to reduce the computational cost and enhance the robustness. But it loses the super performance of maximum likelihood estimation. To reduce the computational cost and have the merit of maximum likelihood estimation, the simultaneous data reconciliation method in an MILP framework is decomposed and replaced by an NMILP subproblem and a quadratic programming (QP) or a least squares estimation (LSE) subproblem. Simulation result of an industrial case shows the high efficiency of the method.

Key words: Data reconciliation, Detection of gross errors, Mixed integer linear programming (MILP), Novel MILP (NMILP) Quadratic programming (QP)

doi:10.1631/jzus.2007.A0904 **Document code:** A **CLC number:** TQ021.8

INTRODUCTION

In industrial process, instrument readings do not meet the laws of conversation and one has to perform data reconciliation to obtain variable estimates. Unfortunately, measured process variables often systematically deviate from their true values. Miscalibrated and malfunctioning instruments are two reasons for biased measurements which are called gross errors. If the measurements are adjusted to meet the laws of conversation in the presence of gross errors, all the adjustments are greatly affected by such biases and would not generally be reliable indicators of the state of the process. So gross errors must be detected and either rectified or discarded before data reconciliation.

Detection of gross errors in steady state process industry has received considerable attention. Most technologies for gross error detection rely on statistical hypothesis testing, such as the global test (GT) (Reilly and Carpani, 1963), the measurement test (MT) (Mah and Tamhane, 1982), the nodal test (NT) (Reilly and Carpani, 1963; Mah *et al.*, 1976), the modified iterative measurement test (MIMT) (Heenan and Serth, 1986), the principal component test (PCT) (Tong and Crowe, 1995), and the general likelihood ratio method (GLR) (Narasimhan and Mah, 1987). These tests have been applied to both measurement and constraint residuals. They are suitable for the detection of one gross error. When multiple gross errors exist, strategies are required to identify them, such as serial elimination, etc.

Yamamura *et al.*(1988) used a model based on Akaike information criterion (AIC) to identify biased measurements. Due to the combinational nature of the problem attempted, a branch-and-bound method was suggested to solve this problem. The combinational

* Project supported by the National Creative Research Groups Science Foundation of China (No. 60421002), and the National "Tenth Five-Year" Science and Technology Research Program of China (No. 2004BA204B08)

algorithm for data reconciliation can be automated by mixed integer linear programming (MILP) techniques (Soderstrom *et al.*, 2001). And the quadratic term of the objective function in the method designed by Yamamura can be reformulated as a linear term in MILP model. But MILP model is not a maximum likelihood estimator for measured variables according to AIC.

To simplify MILP model and achieve the merit of maximum likelihood estimation, two independent models are used in this paper for detection of gross errors and data reconciliation, consistent with AIC in nature. The novel MILP (NMILP) reduces computational cost and has the merit of AIC. Simulation study of an industrial case shows the good performance of the presented method.

PRINCIPLES OF MIXED INTEGER PROGRAMMING FOR DATA RECONCILIATION

Akaike information criterion (AIC)

Data reconciliation and gross error detection can be addressed as a model discrimination and parameter estimation problem, where multiple models correspond to the partitioning of random and gross errors. If more than one of these models can be fitted to the data under consideration, it becomes necessary to identify which model to be used. So, one is interested in obtaining the most likely model and its parameters. Since maximum likelihood estimators are asymptotically efficient under certain condition, the likelihood function is a very sensitive criterion of deviation model parameters from their true values. AIC is an estimate of the Kullback-Leibler mean information for distance between the true model and the model under consideration (Yamamura *et al.*, 1988; Arora and Biegler, 2001). It is given by

AIC=−2log(maximum likelihood)+2(No. of independently adjusted parameters within the model), (1)

and can be rewritten as

$$AIC = E(S) = -2\sum_{i=1}^{N}\log\{l[\varepsilon(i,p)]\} + 2\dim(p), \quad (2)$$

where $E(\cdot)$ is the expectation, ε the measurement error obtained after reconciliation, i the observation, $l(\cdot)$ the likelihood function, and p the number of independently adjusted model parameters. For data reconciliation, we consider the total number of parameters to be given by

$$\dim(p) = \dim(p_0) + n_{\text{out}}, \quad (3)$$

where p_0 is the number of model parameters and n_{out} the number of outliers. Here variables with outlying measurements are treated as parameters because their reconciled values are adjusted only from measurements without gross errors. In this paper, we consider the likelihood function to be a least squares function formed after removing gross errors. We shall observe later that a novel method is designed for efficient detection of gross errors and data reconciliation consistent with the idea of AIC.

Mixed integer approach for data reconciliation

Yamamura *et al.*(1988) used AIC for data reconciliation and parameter estimation for a linear system. The set of measurement sensors (J) were divided into faulty (F) and non-faulty (J–F) sets. For the faulty sensors, they estimated the biases in the following format

$$\min_{\theta_i, v_j}\left\{\frac{1}{2}\sum_{j\in J-F}\theta_j^2 + \frac{1}{2}\sum_{j\in F}(\theta_j - v_j)^2\right\} + |F|, \quad (4)$$

$$\text{s.t.} \quad \sum_{j\in J} a_{ij}\theta_j = b_i, \quad i\in J, \quad (5)$$

where θ_j is the studentized residual, v_j the biases scaled by instrument standard deviations and J the set of equations resulting from eliminations of the measured variables. To systematically select F, they devised a branch-and-bound strategy to select a set of biased sensors and solve Eq.(4). This constituted the branching operation. For bounding the objective functions, the power set of J is divided into two non-intersecting subsets constituting faulty and non-faulty instruments. The procedure can be easily translated into a mixed integer non-linear programming (MINLP) with binary variables identifying faulty sensors. Using a linear model of the measured variables, we state MINLP (Arora and Biegler, 2001) as

$$\min_{x_i, y_i, \mu_i}\sum_{i=1}^{n}\left[\frac{(x_i^{\text{M}} - x_i)}{\sigma_i} - \frac{\mu_i}{\sigma_i}\right]^2 + 2\sum_{i=1}^{n} y_i, \quad (6)$$

$$\text{s.t.}\quad \boldsymbol{Ax}=0, \tag{7}$$
$$|\mu_i|\le U_i y_i, \tag{8}$$
$$|\mu_i|\le L_i y_i, \tag{9}$$
$$y_i\in\{0, 1\}, \tag{10}$$
$$x_i\ge 0, \tag{11}$$

where subscript i refers to the ith variable of the parameters, n is the number of measured variables, x_i^{M} the measurement, x_i the reconciled value, σ_i the standard deviation, y_i a binary variable denoting the existence of bias, $\boldsymbol{A}$ the matrix for constraints, μ_i the magnitude of bias, L_i and U_i the lower and upper bounds on bias respectively. If we use an MINLP solver such as LINGO® to solve Eq.(6), we cannot guarantee all the intermediate mixed integer linear programs (MILPs) be feasible and bounded. To ensure boundedness of MILPs, we add bound constraints of the form

$$x_i<\chi_i. \tag{12}$$

Soderstrom *et al.*(2001) devised an MILP approach to minimize an objective function similar to AIC. The advantage of MILP is that it eliminates the non-linear programming subproblem associated with MINLP algorithm. The quadratic term in the objective Eq.(6) is replaced by the l_1 norm and penalty is added.

$$\min_{x_i, y_i, \mu_i} \sum_{i=1}^{n}\left|\frac{(x_i^{\mathrm{M}}-x_i)}{\sigma_i}-\frac{\mu_i}{\sigma_i}\right|+2\sum_{i=1}^{n}\omega_i y_i, \tag{13}$$
$$\text{s.t.}\quad \boldsymbol{Ax}=0, \tag{14}$$
$$\mu_i\le U_i y_i, \tag{15}$$
$$-\mu_i\le L_i y_i, \tag{16}$$
$$\mu_i-z_iU_i-z_iL_i+L_iy_i\le 0, \tag{17}$$
$$-\mu_i-z_iU_i-z_iL_i+L_iy_i\le L_i+U_i, \tag{18}$$
$$z_i\le y_i, \tag{19}$$
$$y_i, z_i\in\{0, 1\}, \tag{20}$$
$$0\le x_i\le \chi_i, \tag{21}$$

where ω_i is the 'weight' function that penalizes identification of too many biases, z_i a binary variable for the sign of bias value μ_i. The non-differentiability caused by the l_1 norm is removed by rewriting the argument of the absolute value as the difference of two positive numbers:

$$x_i^{\mathrm{M}}-(x_i+\mu_i)=r_i-q_i, \tag{22}$$
$$q_i, r_i\ge 0. \tag{23}$$

To eliminate the absolute value operator, the first term of Expression (13) can be written as

$$x_i^{\mathrm{M}}-(x_i+\mu_i)=r_i+q_i. \tag{24}$$

The objective Expression (13) contains a more robust objective function but does not directly minimize AIC. Also, the choice of the weight functions may be arbitrary, and unsuitable choice of the magnitude ω_i may degenerate MILP performance. So how to design ω_i needs to be further researched. But AIC is a more complete measure of model fitting as it also includes the maximum likelihood on good data.

DETECTION OF GROSS ERRORS AND DATA RECONCILIATION

Both MINLP and MILP are suitable for problems with only a few variables to be reconciled. For large-scale industrial data reconciliation problems, the combinatorial overhead is too great to justify their use, especially online. Also if there are constraints in data reconciliation problems, the computational overhead on MINLP can be large (Arora and Biegler, 2001). To simplify MINLP and MILP, we decompose the problem into two subproblems for detection of gross errors and data reconciliation.

NMILP for gross error detection

To detect gross errors is to locate them. To avoid huge computation in solving the MINLP problem, an NMILP model is designed to detect gross errors only. This method is based on the assumption that the smallest number of gross errors located in a system by an algorithm is the number with the highest probability equal to the true number presented in a system. And that is consistent with the idea of maximum likelihood estimation. The NMILP model can be written as

$$\min\sum_{i=1}^{n} y_i, \tag{25}$$
$$\text{s.t.}\quad \boldsymbol{Ax}=0, \tag{26}$$
$$|x_i^{\mathrm{M}}-x_i-\mu_i|\le\sigma_i, \tag{27}$$
$$|\mu_i|\le U_i y_i, \tag{28}$$
$$|\mu_i|\le L_i y_i, \tag{29}$$
$$y_i\in\{0, 1\}, \tag{30}$$
$$x_i\ge 0. \tag{31}$$

To reduce the computational cost, the constraint Eq.(27) is designed to represent the constraint of μ_i, which replaces the constraint Eq.(22). But, the values of x_i and μ_i are not estimated by maximum likelihood estimators in the proposed algorithm. So the algorithm can only be used to denote the existence of bias in the ith variable. Compared with Soderstrom's MILP, the NMILP model reduces immediate variables such as q_i and r_i, and avoids choosing the weight function ω_i.

In nature, the NMILP model can be deduced from MILP. Since r_i and q_i in Eq.(24) do not have any constraints besides Eqs.(22) and (23), they should be equal to zero to achieve the minimum value of the objective Expression (13). So the MILP model can be reformed easily as an NMILP model.

Data reconciliation and gross error estimation

This subsection considers the problem of data reconciliation. Since the positions of variables with gross errors have been determined, the objection Eq.(6) and its constraints can be formulated as follows

$$\min \sum_{i\in J-F}\left[\frac{x_i^{\mathrm{M}}-x_i}{\sigma_i}\right]^2+\sum_{i\in F}\left[\frac{x_i^{\mathrm{M}}-x_i-\mu_i}{\sigma_i}\right]^2, \quad (32)$$

$$\text{s.t.}\quad \boldsymbol{Ax}=0, \quad (33)$$

$$x_i\geq 0. \quad (34)$$

Here, μ_i can be considered as unmeasured variables to be eliminated. Then the objective Eq.(32) and its constraints can be written as follows

$$\min \sum_{i\in J}\left[(x_i^{\mathrm{M}}-x_i)/\sigma_i\right]^2, \quad (35)$$

$$\text{s.t.}\quad \boldsymbol{Ax}+\boldsymbol{Bu}=0, \quad (36)$$

$$x_i\geq 0. \quad (37)$$

where $\boldsymbol{u}$ is the vector of biases μ_i, $\boldsymbol{A}$ and $\boldsymbol{B}$ are constant matrices. This is a problem of least squares estimation or quadratic programming. In Eq.(36), $\boldsymbol{u}$ can be eliminated by using matrix projection (Crowe *et al.*, 1983).

CASE STUDY

Simulation procedure of a classic case is applied to study the performance of the proposed method. In this paper, MILP and NMILP techniques are evaluated using a Monte Carlo simulation. Each result is based on 1 000 simulation trials where the random errors and the criterion used to judge the performance are the average number of type I errors (AVTI) (Narasimhan and Mah, 1987), the overall power (OP) (Narasimhan and Mah, 1987) and expected fraction of perfect identification (OPF) (Rollins and Davis, 1992). They are defined as follows:

$$AVTI=\frac{\text{No. of unbiased variables wrongly identified}}{\text{No. of simulation trials}}, \quad (38)$$

$$OP=\frac{\text{No. of unbiased variables correctly identified}}{\text{No. of biased variables simulated}}, \quad (39)$$

$$OPF=\frac{\text{No. of trials with perfect identification}}{\text{No. of simulation trials}}. \quad (40)$$

A schematic diagram of recycle process network is shown in Fig.1 (Narasimhan and Mah, 1987). And the true flow rates are shown in Table 1. Each measurement value for simulation trial is taken as the sum of the true value and a random value between $-0.025x_l$ and $0.025x_l$, where x_l is the true value. L_i and U_i are chosen as 0.6 and 4 respectively. Twenty percent of the true value is added to the biased stream to evaluate the proposed method and compare with MILP method. All simulations in this work are performed using an application developed by Visual C++ 6.0. The MILPs and NMILPs are solved using calls to a library of subroutines packages with LINGO® optimization software.

Table 2 and Table 3 show that the NMILP method performs as well as MILP. Under some conditions, such as rows 5~7 in Table 2 and rows 1, 2, 4~10, 12, 13, 20, and 21 in Table 3, NMILP performs better than MILP. The bad performance in Table 3, such as rows 18 and 19, can be explained by the theory of equivalent set (Bagajewicz and Jiang, 1998).

NMILP and MILP avoid gross errors spreading to correct measured values, because they need not use values estimated by the least squares method to detect gross errors. Compared with Table 4 in (Sanchez *et al.*, 1999), we can find that both NMILP and MILP have better performance than the simultaneous estimation of gross errors (SEGE), the unbiased estimation technique (UBET), and the generalized likelihood ratio (GLR).

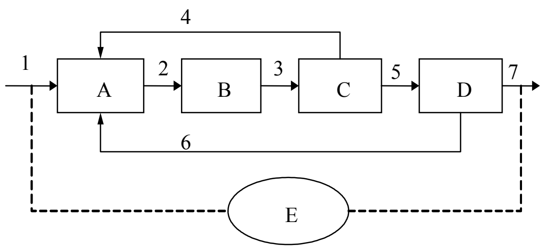

Fig.1 Recycle process network

Table 1 True values for the measured flow rates

l	x_l	l	x_l
1	5	5	10
2	15	6	5
3	15	7	5
4	5		

l: Measurement variables number; x_l: true flow rate

Table 2 Performance of NMILP and MILP when one bias is introduced

No.	Biased stream	NMILP			MILP		
		AVTI	OP	OPF	AVTI	OP	OPF
1	1	0.000	1.000	1.000	0.000	1.000	1.000
2	2	0.000	1.000	1.000	0.000	1.000	1.000
3	3	0.000	1.000	1.000	0.000	1.000	1.000
4	4	0.025	0.958	0.958	0.002	0.993	0.993
5	5	0.000	1.000	1.000	0.002	1.000	0.998
6	6	0.000	1.000	1.000	0.002	1.000	0.998
7	7	0.000	1.000	1.000	0.004	1.000	0.996

Table 3 Performance of NMILP and MILP when two biases are introduced

No.	Biased stream	NMILP			MILP		
		AVTI	OP	OPF	AVTI	OP	OPF
1	1-2	0.000	1.000	1.000	0.001	1.000	0.999
2	1-3	0.000	1.000	1.000	0.003	1.000	0.997
3	1-4	0.024	0.972	0.943	0.003	0.995	0.990
4	1-5	0.000	1.000	1.000	0.007	1.000	0.993
5	2-5	0.000	1.000	1.000	0.000	1.000	1.000
6	2-6	0.000	1.000	1.000	0.002	1.000	1.000
7	2-7	0.000	1.000	1.000	0.001	1.000	0.999
8	3-5	0.000	1.000	1.000	0.000	1.000	1.000
9	3-6	0.000	1.000	1.000	0.000	1.000	1.000
10	3-7	0.000	1.000	1.000	0.003	1.000	0.997
11	4-7	0.002	0.974	0.948	0.001	0.998	0.995
12	5-7	0.000	1.000	1.000	0.008	1.000	0.992
13	1-6	0.035	0.983	0.963	0.216	0.893	0.748
14	1-7	1.000	0.000	0.000	1.004	0.000	0.000
15	2-3	0.000	1.000	1.000	0.000	1.000	1.000
16	2-4	0.056	0.924	0.848	0.000	0.926	0.852
17	3-4	0.068	0.921	0.841	0.002	0.922	0.843
18	4-5	0.415	0.775	0.550	0.066	0.966	0.931
19	4-6	0.197	0.874	0.248	0.070	0.938	0.876
20	5-6	0.737	0.616	0.230	0.952	0.528	0.045
21	6-7	0.035	0.983	0.965	0.218	0.884	0.762

CONCLUSION

A novel mixed integer linear programming model is presented to detect gross errors in industrial process, which is simpler than MILP approach. Simulation showed that NMILP performs better than MILP. Investigations indicated that NMILP method is robust with respect to gross errors and can be easily extended to nonlinear or dynamic systems using integer nonlinear programming method.

References

Arora, A.L., Biegler, L.T., 2001. Redescending estimators for data reconciliation and parameter estimation. *Computers and Chem. Eng.*, **25**:1585-1599. [doi:10.1016/S0098-1354(01)00721-9]

Bagajewicz, M., Jiang, Q., 1998. Gross error modeling and detection in plant linear dynamic reconciliation. *Computers and Chem. Eng.*, **22**(12):1789-1810. [doi:10.1016/S0098-1354(98)00248-8]

Crowe, C.M., Garcia Campos, Y.A., Hrymak, A., 1983. Reconciliation of process flow rates by matrix projection. Part I: linear case. *Am. Inst. Chem. Eng. J.*, **29**(6):881-888.

Heenan, W.A., Serth, R.W., 1986. Gross errors detection and data reconciliation in steam-metering system. *Am. Inst. Chem. Eng. J.*, **32**:733-742.

Mah, R.S.H., Stanley, G., Downing, D., 1976. Reconciliation and rectification of process flow and inventory data. *Ind. Eng. Chem. Process Design Dev.*, **15**:175-183. [doi:10.1021/i260057a030]

Mah, R.S.H., Tamhane, A.C., 1982. Detection of gross errors in process data. *Am. Inst. Chem. Eng. J.*, **28**:828-830.

Narasimhan, S., Mah, R., 1987. Generalized likelihood ratio method for gross error identification. *Am. Inst. Chem. Eng. J.*, **33**:1514-1521.

Reilly, P., Carpani, R., 1963. Application of Statistical Theory of Adjustments to Material Balances. Proc. 13th Can. Chem. Eng. Conf. Montreal, Quebec.

Rollins, D., Davis, J., 1992. Unbiased estimation of gross errors in process measurements. *Am. Inst. Chem. Eng. J.*, **38**:563-572.

Sanchez, M., Romagnoli, J., Jiang, Q., Bagajewicz, M., 1999. Simultaneous estimation of biases and leaks in process plants. *Computers and Chem. Eng.*, **23**:841-857. [doi:10.1016/S0098-1354(99)00104-0]

Soderstrom, T.A., Himmelblau, D.M., Edgar, T.F., 2001. A mixed integer optimization approach for simultaneous data reconciliation and identification of measurement bias. *Control Eng. Practice*, **9**:869-876. [doi:10.1016/S0967-0661(01)00056-9]

Tong, H., Crowe, C.M., 1995. Detection of gross errors in data reconciliation by principal component analysis. *Am. Inst. Chem. Eng. J.*, **41**:1712-1722.

Yamamura, K., Nakajima, M., Matsuyama, H., 1988. Detection of gross errors in process data using mass and energy balances. *Int. Chem. Eng.*, **28**(1):91-98.