Detecting Gross Errors for Steady State Systems

Congli Mei

Jiangsu University, Zhenjiang JS 212013, China
clmei@ujs.edu.cn

Abstract. Gross error detection is important to data reconciliation in process industry. In practice, gross errors cannot be identified exactly by any algorithm. The issue of unreasonable solutions of gross error detection algorithms is discussed. A novel mixed integer optimization method presented in a previous paper is used in this paper. A strategy is proposed to identify gross errors and its most possible alternatives for steady state systems by the method. Gross errors are identified without the need for measurements elimination. The computation results show the effectiveness of the proposed strategy.

Keywords: data reconciliation, gross error detection, equivalent set, mixed integer optimization.

1 Introduction

In industrial process, instrument readings do not satisfy 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. There are two fundamental problems associated with data reconciliation: gross error detection and identification. Meaningful data adjustments can be obtained if and only if there is no gross error in data.

Chemical process variables can be classified into two groups–redundant variables and non-redundant variables. If a variable can be measured directly and estimated indirectly from other relationships(such as mass balance, energy balance, and so on), it is termed as redundant variables and can be reconciled. Non-redundant variables can not be estimated except being measured directly and cannot be reconciled [1]. Several redundancy analysis methods have been proposed in the past. Vaclavek and Loucka proposed a reduced balance scheme(RBS). After deleting units having unmeasured feed or product flows and combining two units connected by unmeasuued flow, the resulting measurements are redundant [1]. Kretsovalis and Mah presented how to identify observability using graph theory [2][3]. Crowe provided a classification algorithm using matrix projection method [4]. Madron and Veverka proposed the multiple Gauss-Jordan elimination algorithm for linear models [5]. Zhang et al. presented a redundancy definition method considering networks topology and sensors' precision [6].

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1947–1954, 2009.

For redundant variables, detecting gross errors in steady state systems has received considerable attention. Most technologies for detection of gross errors rely on statistical hypothesis testing, such as the global test (GT) [7], the measurement test (MT) [8], the nodal test (NT) [7] [9], the modified iterative measurement test (MIMT) [10], the principal component test (PCT) [11], and the general likelihood ratio method (GLR) [12], etc. These tests have been applied to both measurement and constraint residuals. But none of these techniques was satisfying. So it is not a sufficient condition for gross error detection that variables are redundant.

In the view of optimization theory, the existing methods of gross error detection just find a solution satisfying the objective function with minimum value. In a recent paper [13], a series of concepts on equivalency theory were presented to explain the uncertainty of gross error detection. That is, two sets of gross errors are equivalent when they have the same effect in data reconciliation. When simulating either one, that leads to the same value of the objective function.

This paper is devoted to the issue of gross error detection for steady state systems and organized as follows: with the property analysis of steady state systems, conditions for exact gross error identification were pointed out. Next, a model of gross error detection is presented. Using the model, a strategy is proposed to look for the most possible alternatives of the results of a gross error detection algorithm. Simulations show the performance of the proposed strategy.

2 Identification of Gross Errors

The basic model of process measurement data is

$$X = \hat{X} + \sigma + \omega \tag{1}$$

where X is a p-dimensional vector of measurements, $\hat{X}$ is a p-dimensional vector of true values, σ is a p-dimensional vector of gross errors, ω is a p-dimensional vector of random errors.

Theorem 1 (Condition for gross error identification). *Consider a steady state system*

$$A\hat{X} = C \tag{2}$$

without considering random errors, gross errors in redundant measurements can be identified exactly if and only if the coefficient matrix A is full rank.

Obviously, the vector σ can be determined if and only if the coefficient matrix A is full rank when we know the measured data X. But in many cases, A is not full rank in process industry. So gross errors can not be identified exactly.

Theorem 1 can be understand intuitively by the theory of equivalency. The equivalent sets of gross errors are theoretically undistinguishable, because two sets of gross errors are equivalent when they have the same effect in data reconciliation [13]. In other words, there exists an equal possibility that the true locations of gross errors are in one of its equivalent sets. Clearly, a gross error set identified by solving a optimization problem is only one of the feasible solutions of the problem and may be unreasonable.

3 A Model for Gross Error Detection

From the view of theorem 1, more than one of gross error sets can be fitted to the data under consideration, it becomes necessary to identify which gross error set to be used. So gross error detection can be addressed as a model discrimination and parameter estimation problem, where multiple models corresponding to the partitioning of random and gross errors. And 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 models parameters from their true values. The Akaike Information Criterion(AIC) is an estimate of the Kullback-Leibler mean information for distance between the true model and the model under consideration [14]. A novel mixed integer linear programming(NMILP) model was presented based on the idea of AIC [15].

Obviously, a set of gross errors with minimum cardinality can be obtained by the NMILP method [15]. It is not necessary to obtain all equivalent gross error sets, but a equivalent set with minimum cardinality maybe useful to reconcile process measurements. The strategy of SICC is based on the same idea that the probability of containing k gross errors is larger than the probability of containing $k+1$ gross errors, which has a principle objective and is based on identifying one candidate for a gross error at a time and solving the data reconciliation model with bias detection using all suspects [13].

To reduce computational burden, an NMILP model is designed only to detect gross errors [15]. This method is based on the assumption that the smallest number of gross errors located in a system by a algorithm is the number with highest probability equal to true number presented in a system. And that is consistent with the idea of the maximum likelihood estimation. The NMILP model can be written as

$$\min \sum_{i=1}^{n} y_i \tag{3}$$

$$s.t. \quad Ax = 0 \tag{4}$$

$$|x_i^M - x_i - \mu_i| \leq \sigma_i \tag{5}$$

$$|\mu_i| \leq U_i y_i \tag{6}$$

$$|\mu_i| \geq L_i y_i \tag{7}$$

$$y_i \in \{0, 1\} \tag{8}$$

$$x_i \geq 0 \tag{9}$$

The NMILP model can be deduced from AIC also [15].

4 Equivalency of Gross Errors

Theorem 1 is helpful to explain why we cannot locate gross errors correctly based on a mathematical objective function. But only graphic features and concepts of equivalent set are performed [13]. And the strategy for searching equivalent sets needs future work. When a set of gross errors with addition of another stream/leak form a loop, they cannot be identified exactly[13]. So a reasonable equivalent set is helpful for better data reconciliation and maintaining instruments. Based on the assumption as the idea of the SICC, a strategy is presented to exploit the most possible gross error equivalent sets by the NMILP method. The detailed steps can be described as follows:

Step 1. Detect gross errors by the NMILP method.

Step 2. Construct a set of candidates (SC) including all the variables flagged by the NMILP method.

Step 3. Select one variable x_g of the SC as a correct measurement and others of the SC are also looked as gross errors. Then, obtain the equivalent set of x_g by solving (3-9).

Step 4. If all the equivalent sets of the SC have been detected, output the result; Otherwise, go to step 3.

5 Case Study

Simulation procedure of a classic case is applied to study the performance of the proposed strategy. For this paper, the solution of NMILP techniques are evaluated using a Monte Carlo simulation study. Each result is based on 1000 simulation trials where the random errors and the criterion used to judge the performance are the average number of type I errors (AVTI) [12], the overall power (OP) [12] and expected fraction of perfect identification (OPF) [16]. They are defined as follows

$$\text{AVTI} = \frac{\text{No. of unbiased variables wrongly identified}}{\text{No. of simulation trials}} \tag{10}$$

$$\text{OP} = \frac{\text{No. of biased variables correctly identified}}{\text{No. of biased variables simulated}} \tag{11}$$

$$\text{OPF} = \frac{\text{No. of trials with perfect identification}}{\text{No. of simulated trials}} \tag{12}$$

A schematic diagram of recycle process network is shown in Fig. 1 [12]. In the case, the true flow rates are shown in Table 1. Each measurement value for simulation trial 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. 20% of the true value is added to the biased stream to evaluate the proposed method. All simulations in this work are performed using an application developed in the Visual C++ 6.0 programming language. The NMILPs are solved using calls to a library of subroutines packages with software.

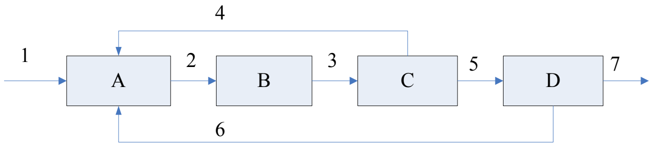

Fig. 1. Recycle process network

Table 1. True values for the measured flow rates

Measurement variables number(l)	True flow rate(x_l)
1	5
2	15
3	15
4	5
5	10
6	5
7	5

Table 2. Performance results for NMILP

No.	Biased stream	NMILP		
		AVTI	OP	OPF
1	1-2	0.000	1.000	1.000
2	1-3	0.000	1.000	1.000
3	1-4	0.024	0.972	0.943
4	1-5	0.000	1.000	1.000
5	2-5	0.000	1.000	1.000
6	2-6	0.000	1.000	1.000
7	2-7	0.000	1.000	1.000
8	3-5	0.000	1.000	1.000
9	3-6	0.000	1.000	1.000
10	3-7	0.000	1.000	1.000
11	4-7	0.002	0.974	0.948
12	5-7	0.000	1.000	1.000
13	1-6	0.035	0.983	0.963
14	1-7	1.000	0.000	0.000
15	2-3	0.000	1.000	1.000
16	2-4	0.056	0.924	0.848
17	3-4	0.068	0.921	0.8413
18	4-5	0.415	0.775	0.550
19	4-6	0.197	0.874	0.248
20	5-6	0.737	0.616	0.230
21	6-7	0.035	0.983	0.965

Table 2 shows the performance of the NMILP method. Some rows of the table have high AVTI and low OP, such as row 14, 18, 19 and 20. These results can be explained with the theorem 1. Totally, the NMILP method avoids gross errors spreading to correct measured values and need't use values estimated by the least squares method to detect gross errors. Compared with the Table 4 in a recent paper[17], we can find that the NMILP has better performance than the simultaneous estimation of gross errors (SEGE), the unbiased estimation technique (UBET), and the generalized likelihood ratio (GLR). So the NMILP method can be used to detect gross errors.

Let us illustrate the proposed strategy to detecting gross errors. Consider the row 20 in Tab 2. Tab 3 shows the procedure of equivalent sets detection by the proposed strategy. The set {4,5} or {4,5,6} obtained by NMILP is a equivalent set of {5,6} and they have the same effect in data reconciliation. But the true location is {5, 6}. It is consistent with the result in the literature [16].

We tested the strategy in variety of situation. In each case the strategy has been successful in identifying the set of gross errors introduced. Table 4 shows the results of introducing biases in steam 2 and 3. Table 5 shows the results of

Table 3. Strategy for exploiting equivalency of gross errors{5,6}

No.	Gross error simulated	Variables flagged by NMILP
Step 1	None	{4, 5}
Step 2	{4,5}	
Step 3	{4 }	{6 }
	{5 }	No feasible solution
Step 4	equivalent sets {4,5},{5,6}, {4,5,6}	

Table 4. Strategy for exploiting equivalency of gross errors{2,3}

No.	Gross error simulated	Variables flagged by NMILP
Step 1	None	{2, 3}
Step 2	{2, 3}	
Step 3	{2 }	No feasible solution
	{3 }	No feasible solution
Step 4	equivalent sets {2,3}	

Table 5. Strategy for exploiting equivalency of gross errors{1,7}

No.	Gross error simulated	Variables flagged by NMILP
Step 1	None	{6 }
Step 2	{6}	
Step 3	{6 }	{1,7 }
Step 4	equivalent sets {6},{1,7}, {1, 6, 7}	

Table 6. Strategy for exploiting equivalency of gross errors{4,5}

No.	Gross error simulated	Variables flagged by NMILP
Step 1	None	{4, 5}
Step 2	{4, 5}	
Step 3	{4 }	{6 }
	{5 }	No feasible solution
Step 4	equivalent sets {4,5},{5,6}, {4,5,6}	

Table 7. Strategy for exploiting equivalency of gross errors{6,7}

No.	Gross error simulated	Variables flagged by NMILP
Step 1	None	{6, 7}
Step 2	{6, 7}	
Step 3	{6 }	{1 }
	{7 }	No feasible solution
Step 4	equivalent sets {6,7},{7,1} {1,6,7}	

introducing biases in stream 1 and 7. Table 6 shows the results of introducing gross errors in stream 4 and 5. Table 7 shows the results of introducing biases in 6 and 7.

6 Conclusion

This paper proposed a strategy to deal with the uncertain results of gross error detection. The nature of the uncertainty is also discussed. In fact, most methods of gross error detection only give the solutions satisfying their objective functions with minimum values. And gross errors cannot be identified exactly by these methods with limited conditions. Using the NMILP method, a strategy for exploiting the most possible gross error set and its most possible alternatives is proposed in the paper. Case study shows the presented strategy possesses good performance.

Acknowledgments

This work was supported by the Natural Science Foundation for Colleges and Universities of Jiangsu Province of China under grant No. 08KJD510011, and the Scientific Research Foundation for Excellent Young Teachers of Jiangsu University of China under grant No. 08JDG017.

References

1. Vaclavek, V., Loucka, M.: Selection of measurements neccessary to achieve multicomponent mass balances in chemical plant. Chemical Engineering Science 31, 1199–1205 (1976)
2. Krestsovalis, A., Mah, R.S.H.: Observability and redundancy classification in generalized process networks. Part I: theorem. Computers and Chemical Engineering 12, 671–687 (1988)
3. Krestsovalis, A., Mah, R.S.H.: Observability and redundancy classification in generalized process networks. Part II: Algorithms. Computers and Chemical Engineering 12, 689–701 (1988)
4. Crowe, C.M., Garcia Campos, Y.A., Hrymak, A.: Reconciliation of process flow rates by matrix projection. Part I: linear case. American institute of chemical engineers Journal 29(6), 881–888 (1983)
5. Madron, F., Veverka, V.: Optimal selection of measuring points in complex plants by linear models. American institute of chemical engineers Journal 38(2), 227–236 (1992)
6. Zhang, P., Rong, G., Wang, Y.: A new method of redundancy analysis in data reconciliation and its application. Computers and Chemical Engineering 25, 941–949 (2001)
7. Reilly, P., Carpani, R.: Applicaiton of statistical theory of adjustments to meterial balances. Proceeding of the 13th Canada chemical engineers Journal 38, 563–572 (1963)
8. Mah, R.S.H., Tamhane, A.C.: Detection of gross errors in process data. American institute of chemical engineers Journal 28, 828–830 (1982)
9. Mah, R.S.H., Stanley, G., Downing, D.: Reconciliation and rectification of process flow and inventory data. Industrial engineering chemical process design development 15, 175–183 (1976)
10. Heenan, W.A., Serth, R.W.: Gross errors detection and data reconciliation in steam-metering system. American institute of chemical engineers Journal 32, 733–742 (1986)
11. Tong, H., Crowe, C.M.: Detection of gross errors in data renciliation by principal component analysis. American institute of chemical engineers Journal 41, 1712–1722 (1995)
12. Narasimhan, S., Mah, R.S.H.: Generalized likelihood ratio method for gross error identification. American institute of chemical engineers Journal 33, 1514–1521 (1987)
13. Bagajewicz, M., Jiang, Q.: Gross error modeling and detection in plant linear dynamic reconciliation. Computers and Chemical Engineering 22(12), 1789–1810 (1998)
14. Yamamura, K., Nakajima, M., Matsuyama, H.: Detection of gross errors in process data using mass and energy balances. International chemical engineering 28(1), 91–98 (1988)
15. Mei, C., Su, H., Chu, J.: Detection of gross errors using mixed integer optimization approach in process indurstry. Journal of Zhejiang University SCIENCE A 8(6), 904–909 (2007)
16. Rollins, D., Davis, J.: Unbiased estimation of gross errors in process measurements. American institute of chemical engineers Journal 38, 563–572 (1992)
17. Sanchez, M., Romagnoli, J., et al.: Simultaneous estimation of biases and leaks in process plants. Computers and Chemical Engineering 23, 841–857 (1999)