15th Triennial World Congress, Barcelona, Spain

STEADY-STATE BILINEAR DATA RECONCILIATION DEALING WITH SCHEDULING

Zhang Puming and Rong Gang[1]

National Key Laboratory of Industrial Control Technology,
Zhejiang University, Hangzhou, P. R. China, 310027

Abstract: Mass balance model for steady-state linear data reconciliation is insufficient, especially when short-term scheduling is performed on hybrid systems. Scheduling-equations are established and used as addition to the model. A new formulation of data reconciliation is thus proposed as steady-state bilinear data reconciliation. In this way, the redundancy degree is increased and the solvability of data reconciliation is improved. Simulation results demonstrate the efficiency and consistency of the proposed approach. *Copyright © 2002 IFAC*

Keywords: data handling systems, scheduling, data models, data reconciliation

1. INTRODUCTION

Since measurements of process variables, such as flow rates, concentrations, temperatures and so on, are subject to errors (both random errors and gross errors), it cannot be expected that any set of measurements will obey the laws of conservation. Data reconciliation is a procedure of optimally adjusting measured data so that the adjusted values obey the conservation laws and other constraints (Crow, 1996).

Although no plant operates at a true time-invariant steady state, in practice the steady state is defined to mean constancy of the mean values of measurements over a given period of time, which is called reconciliation period. For refinery, steady-state linear data reconciliation is usually performed on flow rates based on mass balance. These mass balance equations are the model of the measurements network. Data reconciliation's results rely greatly on the redundancy degree of the model (Kretsovalis, 1987). If the model's redundancy degree is very low, it will make some unmeasured variables unobservable and some gross errors uncorrected.

Refineries have not only continuous flows, but also discrete scheduling events in order to meet production and sales requirements. Scheduling, especially short-term changes of plant outlets and oil movements, make it difficult to build mass balance equations. They are typical hybrid systems where conventional steady-state linear data reconciliation is very difficult to be implemented. There are two conventional ways to deal with this problem. One is to modify mass balance equations when the short-term scheduling events happen. This is not a

[1] Author to whom correspondence should be addressed
Phone: 86-571-87951125
Fax: 86-571-87951445
Email: grong@iipc.zju.edu.cn

good way, because the reconciliation period will be modified, which will bring troubles to yield accounting and other applications. The other way is to include all of the short-term scheduling schemes in mass balance equations. For example, the first bypass of ordinary pressure distillation column has three branch ways for solvent, raw material of catalystic reformer and kerosene, respectively. The first bypass product can be either of them according to short-term scheduling schemes. A balance equation can be established to model these scheduling schemes. Since there is only one sensor on the bypass, and no sensors on branch ways, flow rates of solvent, raw material of catalystic reformer and kerosene are all unmeasured. This equation has three unmeasured variables. Thus the model's redundancy degree is reduced and will deteriorate performance of the data reconciliation. As a result, this method is also not very practical.

Actually, a scheduling scheme consists of some useful information such as the scheduling schemes execution time and what the branch way's product is at that time. In this paper, this information is used to build scheduling-equations. Consequently the redundancy degree of the model, including mass balance equations and scheduling-equations, will be improved. Then a new formulation of data reconciliation is proposed as steady-state bilinear data reconciliation. Simulation results demonstrate the efficiency and consistency of the proposed approach.

This paper is organized as follows. Steady-state linear data reconciliation is briefly discussed in Section 2. Section 3 gives out the proposed bilinear data reconciliation dealing with scheduling. Simulation results are presented in Section 4 and Section 5 concludes the whole paper.

2. STEADY-STATE LINEAR DATA RECONCILIATION

The steady-state linear model of a process is usually defined as follows

$$\mathbf{Ax} + \mathbf{Bu} = 0 \tag{1}$$

Where $\mathbf{A}_{n\times p}$ is matrix corresponding to measured variables, $\mathbf{B}_{n\times q}$ is matrix corresponding to unmeasured variables, $\mathbf{x}$ is p-dimensional vector of measured variables, $\mathbf{u}$ is q-dimensional vector of unmeasured variables.

The steady-state linear data reconciliation problem can be written as follows

$$\begin{aligned} &\underset{\hat{x}}{\text{Min}} \quad \tfrac{1}{2}(\hat{\mathbf{x}}-\mathbf{x})^{T}\mathbf{S}_{x,m}^{-1}(\hat{\mathbf{x}}-\mathbf{x}) \\ &\text{s.t.} \qquad \mathbf{A}\hat{\mathbf{x}}+\mathbf{Bu}=0 \end{aligned} \tag{2}$$

Where $\mathbf{x}$ is p-dimensional vector of measurements, $\hat{\mathbf{x}}$ and $\mathbf{u}$ are the reconciled and estimated values vectors, $\mathbf{S}_{x,m}$ is the covariance matrix of $\mathbf{x}$.

Assume gross errors have been detected and treated as unmeasured variables, and random measurement errors are independent and normally distributed with zero mean and known covariance matrix $\mathbf{S}_{x,m}$. Then the steady-state linear data reconciliation problem (2) can be easily solved by using least squares algorithm.

The reconciliation results depend greatly on the model's redundancy degree. If the redundancy degree is too low, it will make some unmeasured variables unobservable and some gross errors uncorrected. Since the linear model includes all of the short-term scheduling schemes, there are too many unmeasured variables and its redundancy degree is very low. Thus, the model Equation (1) should be revised to increase its redundancy degree to guarantee the reconciliation performance.

3. STEADY-STATE BILINEAR DATA RECONCILIATION DEALING WITH SCHEDULING

Assume there are s scheduling schemes during the *j*th reconciliation period, and all of them are only performed on node k, as shown in Fig 1. x_0 is measured, $u_1 \sim u_s$ are all unmeasured and the reconciliation period is T.

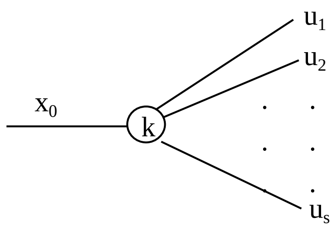

Fig 1 Sketch map of scheduling

Suppose
*j*th reconciliation period beginning time $t_0 = (j-1)T$,
*j*th reconciliation period ending time $t_s = jT$,
from t_0 to t_1: $u_1 = x_0,\ u_2 = u_3 = \ldots = u_s = 0$;
from t_1 to t_2: $u_2 = x_0,\ u_1 = u_3 = \ldots = u_s = 0$;
…
from t_{s-1} to t_s: $u_s = x_0,\ u_1 = u_2 = \ldots = u_{s-1} = 0$;
if $t_{i-1} = t_i$, the *i*th scheduling scheme has not happened.

Mass balance equation of node k is

$$x_0 = \sum_{i=1}^{s} u_i \tag{3}$$

Obviously, this equation includes all s scheduling schemes and there are s unmeasured variables. In fact, when one scheduling scheme is performed, only one branch way has mass flow, and the other s-1 branch ways have no flows. That means actually only one variable is unmeasured variable and the others are zero. So when using Equation (3) as a equation of the model, the redundancy degree is very low.

Let

$$\Delta t_i = t_i - t_{i-1} \quad (i = 1, 2, \ldots, s) \qquad (4)$$

then

$$u_i = \frac{\Delta t_i}{T} x_0 \quad (i = 1, 2, \ldots, s) \qquad (5)$$

$$\sum_{i=1}^{s} \Delta t_i = T \qquad (6)$$

Equation (5) and (6) are defined as scheduling-equations of node k. Obviously the scheduling-equations have close relationship with Equation (3). Although the execution time of scheduling schemes is definitely known, as other measurements, the actually execution time is polluted by errors (both random and gross errors). Suppose in the absent of gross errors, Δt_i (i=1,2, ⋯,s) are measured variables whose covariance matrix is $\boldsymbol{\Sigma_{st}}$. Obviously, Equation (6) is independent and consist of measured variables. A system has a degree of redundancy k, when, at least, k linearly independent balance equations can be written using measured variables only (Bagajewicz and Sanchez, 1999). According to that definition, the redundancy degree will at least increase one if the scheduling-equations are used in the model instead of Equation (3).

Define the data reconciliation problem of *j*th period is

$$\min_{\hat{\mathbf{x}},\Delta\mathbf{t}} \ (\hat{\mathbf{x}} - \mathbf{x})^T \boldsymbol{\Sigma}^{-1} (\hat{\mathbf{x}} - \mathbf{x}) + (\hat{\boldsymbol{\Delta}}\mathbf{t} - \boldsymbol{\Delta}\mathbf{t})^T \boldsymbol{\Sigma}_{\mathbf{st}}^{-1} (\hat{\boldsymbol{\Delta}}\mathbf{t} - \boldsymbol{\Delta}\mathbf{t}) \qquad (7)$$

$$\text{s.t.} \qquad \mathbf{A_1}\hat{\mathbf{x}} + \mathbf{B_1}\mathbf{u} = 0 \qquad (8)$$

$$u_i = \frac{\hat{\Delta t}_i}{T} \hat{x}_0 \quad (i = 1, 2, \ldots, s) \qquad (9)$$

$$\sum_{i=1}^{s} \hat{\Delta t}_i = T \qquad (10)$$

Where $\mathbf{A_1}$, $\mathbf{B_1}$ are obtained by deleting the row vector that correspond to node k in the matrixes $\mathbf{A}$ and $\mathbf{B}$ (to ensure the constraints are independent), $\hat{\boldsymbol{\Delta}}\mathbf{t}$ is the reconciled values of $\boldsymbol{\Delta}\mathbf{t}$.

Obviously that is a steady-state bilinear data reconciliation problem. SQP (Successive Quadratic Programming) algorithm is used to solve this problem in simulation.

In this way, the redundancy degree of the model is increased and the solvability of the data reconciliation problem is improved. Although this paper discusses the case where there are scheduling schemes only on one node, the method is also applicable to cases where scheduling events happen on more than one node.

4. SIMULATION

The simulation network consists of 6 nodes and 13 variables as shown in Fig 2. Among those variables, x_1~x_9 are measured and u_1~u_3 are unmeasured. For w, two cases, unmeasured or measured, are discussed.

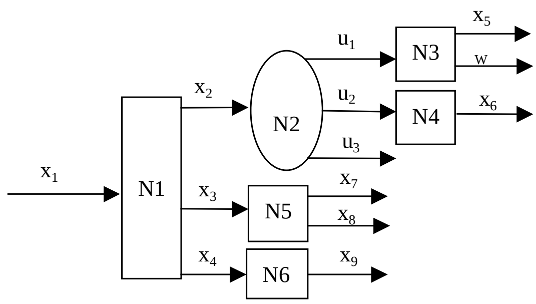

Fig 2 Simulation network

4.1 Redundancy degree

Suppose the reconciliation period T = 24 hours, and scheduling is only performed on node N2:
t = 0~8: Δt_1=8, $u_1 = x_2$, $u_2 = u_3 = 0$;
t = 8~16: Δt_2=8, $u_2 = x_2$, $u_1 = u_3 = 0$;
t = 16~24: Δt_3=8, $u_3 = x_2$, $u_1 = u_2 = 0$.

The mass balance equations are

$$x_1 = x_2 + x_3 + x_4 \qquad (11)$$

$$x_2 = u_1 + u_2 + u_3 \qquad (12)$$

$$u_1 = x_5 + w \qquad (13)$$

$$u_2 = x_6 \qquad (14)$$

$$x_3 = x_7 + x_8 \qquad (15)$$

$$x_4 = x_9 \qquad (16)$$

Whether w is measured or not, the redundancy degree is 3 because only Equations (11), (15) and (16) consist of measured variables.

Using the method proposed in this paper, the scheduling-equations of node N2 are

$$u_1 = \frac{\Delta t_1}{T} x_2 \qquad (17)$$

$$u_2 = \frac{\Delta t_2}{T} x_2 \qquad (18)$$

$$u_3 = \frac{\Delta t_3}{T} x_2 \qquad (19)$$

$$\Delta t_1 + \Delta t_2 + \Delta t_3 = T \qquad (20)$$

While including scheduling-equations (17)~(20) in the model, to guarantee the independence, Equation (12) must be deleted from the model. Thus, the model of the steady-state bilinear data reconciliation problem dealing with scheduling can be stated as follows

$$x_1 = x_2 + x_3 + x_4 \qquad (21)$$

$$u_1 = x_5 + w \qquad (22)$$

$$u_2 = x_6 \qquad (23)$$

$$x_3 = x_7 + x_8 \qquad (24)$$

$$x_4 = x_9 \qquad (25)$$

$$u_1 = \frac{\Delta t_1}{T} x_2 \qquad (26)$$

$$u_2 = \frac{\Delta t_2}{T} x_2 \qquad (27)$$

$$u_3 = \frac{\Delta t_3}{T} x_2 \qquad (28)$$

$$\Delta t_1 + \Delta t_2 + \Delta t_3 = T \qquad (29)$$

When w is unmeasured, combining Equation (23) and (27) can get a new equation that consists of measured variables. Obviously Equation (29) consists of measured variables. Then, there are 5 independent equations consist of measured variables. Thus, the redundancy degree is 5. By using the same analysis, it is easy to know when w is measured, the redundancy degree will be 6. So, the model's redundancy degree is increased when including scheduling-equations in the model.

4.2 Simulation results

Assume there are no gross errors, and the random errors of all the measurements (both flow rates and scheduling execution time) are independent and normally distributed with zero mean and known covariance matrix with their diagonal elements shown in Table 1 and Table 2. The measured values and true values are also shown in Table 1 and Table 2. One thousand samples are generated for Monte Carlo simulation.

Case I: When w is unmeasured, according to Vaclavek's criterion (Vaclavek, 1969), the unmeasured variables u_1, u_3 and w are unobservable, because they form a circle through node N2, N3 and the environment. Consequently, the conventional linear data reconciliation based on the mass balance model can not be performed. However, the proposed bilinear data reconciliation based on the model including both mass balance equations and scheduling-equations can be performed properly. The reconciliation result of one sample is shown in Table 1 and Table 2.

Case II: When w is measured, both of the conventional approach and the proposed approach can be performed. Define the performance index of data reconciliation results (IRR) as follows

$$\mathrm{IRR} = \frac{1}{Ns} \sum_{k=1}^{Ns} \sum_{i=1}^{n} \left(\frac{\hat{\mathbf{x}}_{i,k} - \mathbf{x}_{ti}}{\mathbf{x}_{ti}} \right)^2 \qquad (30)$$

Where Ns is the number of samples, $\hat{\mathbf{x}}_{\mathbf{i,k}}$ is the reconciled value of variable i of the *k*th sample, $\mathbf{x}_{\mathbf{ti}}$ is the true value of variable i.

Comparison result of one sample is shown in Table 3 and the IRR comparison results are shown in Table 4.

Obviously, the proposed approach is much more effective than the conventional approach and its consistency is well shown.

5. CONCLUSIONS

Since mass balance model for conventional steady-state linear data reconciliation is insufficient sometimes, especially when short-term scheduling events are performed in hybrid systems, scheduling-equations are established and used as addition to the model and a new formulation of data reconciliation is proposed as steady-state bilinear data reconciliation. In this way, the redundancy degree is increased and the solvability of data reconciliation is improved. Comparisons between the proposed approach and the conventional approach are made in the simulation. Simulation results demonstrate the efficiency and consistency of the proposed approach. However, the assumption about the statistic feature of scheduling execution time variables is not so proper. This problem will be the focus of our future work.

ACKNOWLEDGMENTS

This work was supported by National High-Tech Research and Development Program of China (No. 2001AA413220 and No. 2001AA411210).

REFERENCES

Bagajewicz, M. J. and M. C. Sanchez (1999). Design and upgrade of nonredundant and redundant linear sensor networks. *AIChE J.*, **45(9**), 1927-1938

Crowe, C. M. (1996). Data reconciliation — progress and challenges. *J. Proc. Cont*., **6(2/3)**, 89-98

Kretsovalis, A. and R. S. H. Mah (1987). Effect of redundancy on estimation accuracy in process data reconciliation. Chem. Engng Sci. **42(9)**, 2115-2121

Vaclavek, V. (1969). Studies on system engineering-III: Optimal choice of the balance measurements in complicated chemical engineering systems. *Chem. Engng Sci*., **24(6)**, 947-955

Table 1 Simulation results of case I (flow rates)

Variables	$\mathbf{x}_t$	$(\Sigma)_{i,i}$	$\mathbf{x}$	$\hat{\mathbf{x}}$
x_1	1000	100	998.83	1001.00
x_2	300	9	299.55	299.20
x_3	300	9	298.08	301.95
x_4	400	16	402.43	399.87
x_5	50	0.25	49.43	49.43
x_6	100	1	99.84	99.89
x_7	100	1	101.04	100.59
x_8	200	4	203.18	201.37
x_9	400	14	397.66	399.87
u_1	100	no	unmeasured	99.93
u_2	100	no	unmeasured	99.89
u_3	100	no	unmeasured	99.38
w	50	no	unmeasured	50.50

Table 2 Simulation results of case I (scheduling executive time)

Variables	$\mathbf{\Delta t}_t$	$(\Sigma_{st})_{i,i}$	$\mathbf{\Delta t}$	$\hat{\mathbf{\Delta t}}$
Δt_1	8	0.09	7.71	8.02
Δt_2	8	0.09	7.76	8.01
Δt_3	8	0.09	7.66	7.97

Table 3 Simulation results of case II

Variables	$\mathbf{x}_t$	$(\Sigma)_{i,i}$	$\mathbf{x}$	$\hat{\mathbf{x}}$ of proposed approach	$\hat{\mathbf{x}}$ of conventional approach
x_1	1000	100	999.74	998.43	996.17
x_2	300	9	295.41	298.25	295.73
x_3	300	9	302.02	300.00	300.07
x_4	400	16	399.61	400.19	400.37
x_5	50	0.25	50.21	50.11	50.21
x_6	100	1	99.96	99.47	99.96
x_7	100	1	100.09	100.33	100.35
x_8	200	4	198.71	199.67	199.73
x_9	400	14	400.55	400.19	400.37
u_1	100	no	unmeasured	100.07	100.28
u_2	100	no	unmeasured	99.47	99.96
u_3	100	no	unmeasured	98.70	95.50
w	50	0.25	50.07	49.96	50.07

Table 4 IRR comparison results of case II

Ns	IRR of proposed approach	IRR of conventional approach
100	0.0014594	0.0019352
1000	0.0014506	0.0018688