Reformulation of Data Reconciliation Problem with Unknown-but-Bounded Errors

José Ragot* and Didier Maquin

Ecole Nationale Supérieure de Géologie, Centre de Recherche en Automatique de Nancy, Institut National Polytechnique de Nancy, 2, Avenue de la Forêt de Haye, F-54516 Vandoeuvre les Nancy

In this paper, a new formulation of the problem of mass and energy balance equilibration in the case of unknown-but-bounded errors is proposed. The bounds of the errors are specified over both a measurement noise and the balance equations. Both bounds are mainly motivated by experimental considerations of the measurement precision; with a more general interpretation, they can be considered as parameters that the user has to adjust to make the reconciliation possible. The method is particularly suitable for linear models but has been extended to nonlinear ones as well. Simulations provide results that compare favorably with those of classical reconciliation methods involving maximum likelihood estimation based on statistical knowledge of the measurement errors.

1. Introduction

The problems of obtaining reliable estimates of process variables from measurements (data validation) and of detecting and isolating gross errors have been well studied. Historically speaking, likely because of measurement availability, static redundancy equations were first utilized in the mineral processing and the chemical industries. The first studies of Ripps (1962),[1] Vaclavek (1969),[2] and Smith and Ichiyen (1973)[3] were concerned with data reconciliation using the now classical technique of equilibration of production balances. In subsequent stages, this data reconciliation principle was generalized to processes that are described by algebraic equations that are either linear in the case of total flow rates[4] or nonlinear in the case of chemical concentrations.[5,6]

At the same time, data reconciliation was employed for more general applications than establishing statistically coherent balances. It was then applied to more fundamental problems such as detection, localization, and estimation of gross errors;[7,8] diagnosis and observability of systems;[9,10] optimization of sensor locations;[11,12] and the study of the reliability of a measurement system.[13] Specific work concerning the numerical regularization of the estimation technique using a projection matrix has been performed by Kelly.[14]

Most approaches to data reconciliation ensure that the estimates of process variables satisfy the material and energy balances in either their linear or nonlinear form. The second important point is that reconciliation is possible only if redundancy equations, i.e., equations containing only redundant variables, are available. Recall that a measured variable is called redundant if it can be calculated from the remaining measured variables. As previously mentioned, this redundancy generally leads to a discrepancy between the equations and the data that have to be reconciled; thus, it provides a check on the reliability of a given set of measurements. Last, it seems relevant to validate and adjust the measurements, taking into account the degree of precision of each measurement and key physical laws. Most of the methods for this use techniques based on statistical considerations, where the noise affecting the records is often characterized by the mean and the covariance of an amplitude probability function. Maximizing the likelihood function resulting from this probability function allows one to express the estimation of the true data. A survey of the methods used in data reconciliation can be found in Crowe.[15] In 1996, Crowe[16] investigated another formulation of the problem of data reconciliation by using the concept of information entropy; this approach allows one to deduce the probability distributions of the data by taking into account the bounds on the data and/or the variance−covariance matrix of the data. For the interested reader, some books on synthesis concerning the subject of data reconciliation are available, including those of Ragot et al.,[17] Romagnoli and Sanchez,[18] Bagajewicz,[19] and Narasimhan and Jordache.[20]

As mentioned above, most previous investigators have used statistical criteria such as least squares as the criteria for calculating the best estimates of process variables. However, the results provided by such estimators are valid only under the following restrictive conditions: (1) the nature of the measurement noise must be known and (2) the model of the process must be known perfectly.

It can be a difficult task to estimate the validity of such assumptions, and in a certain number of applications, it is obvious that both conditions are not fully satisfied. Therefore, it becomes very hazardous and mathematically incorrect to reconcile operation data with regard to an uncertain model without taking this fact into account. Some attempts in this direction have been already published by Mandel et al.[21] and Maquin et al.[22] An alternative method is proposed here that does not use any hypothesis about the noise distribution. The only information needed about the noise is the value of its bounds. As the process model might be inaccurate, it is not necessary that the estimate exactly verify the model; it is only desired that the model residual belong to a given interval. Indeed, our strategy is based on interval constraint satisfaction both for the variable estimations and for the model residuals. From a historical point of view, the 1985 paper of Himmelblau[23] is

* To whom correspondence should be addressed. Tel: 33 + 383-59-56-82. Fax: 33 + 383-59-56-44. E-mail: jragot@ensem.inpl-nancy.fr; Jose.Ragot@ensg.inpl-nancy.fr.

10.1021/ie020213s CCC: $27.50
Published on Web 02/11/2004

probably one of the first to give a formulation of such data rectification principles. This idea was confirmed in the works of Kyriakopoulos and Kalitventzeff[24] and Harikumar and Narasimhan.[25] In fact, the proposed method was based on a least-squares estimate subject to inequality constraints. More recently, Dovi and Del Borghi[26] generalized the fundamental method of data reconciliation developed by Crowe[16] to allow for the presence of censored measurements. In a paper published in 2001, Soderstrom et al.[27] suggested a reconciliation method that also uses inequality constraints for the determination of the bias within a mixed integer optimization framework.

Here, our strategy involves inequality constraints on the estimates and on the residuals of the balance equations and can be considered as an alternative of this idea. Moreover, the probability distribution of the measurement errors is not used, and only bounds on these errors are considered, which is, perhaps, a less restrictive approach. The aim of this presentation, based on the previous paper of Mandel et al.,[21] is to provide the elements necessary for the implementation of a procedure using this type of formalization based on inequality constraints.

In section 2, we recall the principle of reconciliation of measures based on a model and a measurement system, and we give an alternative approach to solve this classical problem based on the satisfaction of inequality constraints set by the user. Then, in section 3, we propose an extension of the technique to nonlinear systems. Section 4 provides numerical results and a discussion.

2. Data Reconciliation. The Linear Case

A linear system (under steady-state conditions) can be described, in the fault-free case, by the following relations

model equation:

$$Mx^* = 0,\ x^* \in \mathscr{R}^n,\ M \in \mathscr{R}^{m\times n} \quad (1a)$$

measurement equation:

$$z = x^* + \epsilon,\ z \in \mathscr{R}^n \quad (1b)$$

where x^* is the vector of process variables, z is the vector of measurements, M is the $m \times n$ matrix of model equations (without loss of generality, M is assumed to be of full row rank), and ϵ is a vector of errors due to measurement noise.

The estimation or data reconciliation problem of system 1 involves finding a set of adjustments such that the adjusted values verify model eq 1a with a given level of satisfaction.

In the context of unknown-but-bounded error, the noise ϵ is assumed to verify a relation of the form

$$|\epsilon| \le e \quad (2a)$$

where e is a known bound. Consequently, the estimations $\hat{x}$ are subjected to the constraints

$$|z - \hat{x}| \le e \quad (2b)$$

In the context of an approximate model, bounds on the balance equations can be imposed; thus, we are looking to adjust the estimates $\hat{x}$ such that

$$|M\hat{x}| \le r \quad (2c)$$

where the bound r is chosen by the user and can also be fixed to zero if the user wishes to verify the balance equations exactly. In this alternative data reconciliation approach, it is thus desired to correct the raw data with bounds (eq 2b) and to satisfy approximately the model equation with a bounded residual (eq 2c). We can say that, for the user, this is a natural way to reconcile the data. Indeed, the user has certain knowledge about the quality of the measurements; the user has also information concerning the precision of the model. Consequently, he or she can define a maximum value for the correction of each variable and a kind of satisfaction level for the process equations. It should be noticed that the first point of view has already been taken into account in the literature; see, for example, the papers of Harikumar and Narasimhan[25] and Narasimhan and Harikumar.[28]

Thus, summarizing, the problem of data reconciliation can be turned into the following constraints

$$\begin{aligned} |z - \hat{x}| &\le e \\ |M\hat{x}| &\le r \end{aligned} \quad (3)$$

In a more sophisticated situation, the bounds can be expanded according to upper and lower bounds corresponding to asymmetrical corrections and satisfaction levels of the constraints

$$\begin{aligned} \underline{e} \le z - \hat{x} \le \bar{e} \\ \underline{r} \le M\hat{x} \le \bar{r} \end{aligned} \quad (4)$$

For numerical solution, the system 4 is expressed using only simple inequalities, i.e.

$$\begin{aligned} &-\hat{x} + z - \bar{e} \le 0 \\ &\hat{x} - z + \underline{e} \le 0 \\ &M\hat{x} - \bar{r} \le 0 \\ &-M\hat{x} + \underline{r} \le 0 \end{aligned}$$

or

$$\begin{pmatrix} -I \\ I \\ M \\ -M \end{pmatrix}\hat{x} + \begin{pmatrix} z - \bar{e} \\ -z + \underline{e} \\ -\bar{r} \\ \underline{r} \end{pmatrix} \le 0 \quad (5)$$

The bounds $\underline{e}$ and $\bar{e}$ are chosen as functions of empirical knowledge concerning the state of the process and, in particular, the probable domains of variation of the different variables. As these bounds express the amplitudes of the corrections applied to the measurements, they can be indexed on an estimation of the precision of these measurements. The bounds $\underline{r}$ and $\bar{r}$ are linked to the degree of satisfaction of the balance constraints and depend on the relative importance given to the different balance equations.

Thus, the data reconciliation problem is to find estimates that satisfy constraints 5. It should be noted that, when the bound r is set to zero, the balance equation is forced to be perfectly satisfied. Because of the nature of the problem (i.e., the presence of inequality constraints), it is clear that the estimates can no longer be obtained analytically. Moreover, there is no guarantee that the estimates will be unbiased or will have the minimum variance. However, many tools are available to solve a system of linear inequalities, all of which are referred to as the LMI (linear matrix inequality) concept;[29] the reader can refer to a summary explanation of the solution of LMIs provided in the Appendix.

Generally speaking, for technical and/or financial reasons, a process might not be completely instrumented, in which case some variables are not measured. This point can easily be taken into account, in the problem of balance formulation, by means of a matrix C (of dimension $p \times n$) of measurement selection. The observability analysis that is required by the presence of unmeasured variables is not explained here, but the reader can consult ref 8 for more details. The problem of the reconciliation of measurements is then expressed by the set of inequalities

$$\begin{aligned} &-C\hat{x} + z - \bar{e} \leq 0 \\ &-z + C\hat{x} + \underline{e} \leq 0 \\ &M\hat{x} - \bar{r} \leq 0 \\ &-M\hat{x} + \underline{r} \leq 0 \end{aligned} \tag{6}$$

where $z \in \mathscr{R}^p$, $p \leq n$. Indeed, LMI system 6 is a straightforward generalization of system 5.

For the problem in which all variables are measured (system 5), a reduction of the number of inequality constraints is observed, as corrective terms can be calculated for measured variables only. Another way of proceeding consists of keeping the set of equations 5 and relaxing the constraints related to the unmeasured variables. Thus, for instance, if the variable k is not measured, a "large" threshold is chosen for the element of row k in the vector $\bar{e}$, a "small" threshold is chosen for the element of same row of vector $\underline{e}$, and the missing measurement can be replaced arbitrarily by 0; from a practical point of view, this is equivalent to not enforcing any proximity between the kth variable and its measurement (since the latter does not exist!).

3. Data Reconciliation. The Nonlinear Case

We now consider the more general case where both the model and measurement systems are nonlinear and described by the expressions

$$\begin{aligned} &f(x^*) = 0, \; f \in \mathscr{R}^m \\ &z = h(x^*) + \epsilon, \; h \in \mathscr{R}^p \end{aligned} \tag{7}$$

This general description includes the case where the device measurement is linear, i.e., where h reduces to identity. Thus, according to the previous strategy, the data reconciliation problem is to find estimates $\hat{x}$ that satisfy the bounds on process variables. Thus, generalizing system 6, the state estimation has to verify the inequality constraints

$$\begin{aligned} &-h(\hat{x}) + z - \bar{e} \leq 0 \\ &h(\hat{x}) - z + \underline{e} \leq 0 \\ &f(\hat{x}) - \bar{r} \leq 0 \\ &-f(\hat{x}) + \underline{r} \leq 0 \end{aligned} \tag{8}$$

By applying the principle of linear balance reconciliation again, we propose linearizing the equations of balances and using the LMI technique to solve the system thus produced. At the outset, to apply this algorithm, it is necessary to have an initial estimation, $\hat{x}^{(0)}$, of the variables; for the measured variables, the measurements provide appropriate initial values, and for the unmeasured variables, the user will be guided in his choice by the a priori knowledge he or she might have of the process. At step i, linearization of the constraint functions f and h around a previous solution $\hat{x}_i$ gives

$$\begin{aligned} &\tilde{f}(x) = f_i + F_i x \\ &\tilde{h}(x) = h_i + H_i x \end{aligned} \tag{9a}$$

with the definitions

$$\begin{aligned} &F_i = \left.\frac{\partial f}{\partial x^{\mathrm{T}}}\right|_{x=\hat{x}_i} \\ &f_i = f(\hat{x}_i) - F_i\hat{x}_i \\ &H_i = \left.\frac{\partial h}{\partial x^{\mathrm{T}}}\right|_{x=\hat{x}_i} \\ &h_i = h(\hat{x}_i) - H_i\hat{x}_i \end{aligned} \tag{9b}$$

The data reconciliation problem defined by eqs 8 can be replaced by the problem

$$\begin{aligned} &-H_i\hat{x} - h_i + z - \bar{e} \leq 0 \\ &H_i\hat{x} + h_i - z + \underline{e} \leq 0 \\ &F_i\hat{x} + f_i - \bar{r} \leq 0 \\ &-F_i\hat{x} - f_i + \underline{r} \leq 0 \end{aligned} \tag{10}$$

which corresponds to the analysis of a set of linear inequalities.

Thus, an iterative algorithm can be easily constructed for solving the nonlinear data reconciliation problem as follows:

E1: Set $i = 0$. Select an initial value for $\hat{x}_i$ (for the measured variables, their corresponding values can be used as initial values for the estimates). If measurements are given with an interval representation, an initial value can be chosen belonging to this interval.

E2: Compute the gradients of the functions f and h (eqs 9b), and linearize the model equation around solution at step i (eqs 9a)

E3: Collect all of the constraints of the estimation problem (problem 10)

E4: Using an LMI routine, solve the linear matrix inequations, and set $\hat{x}_{i+1} = \hat{x}$

E5: Test for convergence of the solution by analyzing the series $\hat{x}$. If $\hat{x}$ is acceptable, then stop the procedure; otherwise, $i \rightarrow i + 1$ and go to step E2.

4. Examples

First Example. Consider the following academic model depending on six variables (with $a = 5.5254$)

$$\begin{aligned} &x_1x_2 + x_5 - a = 0 \\ &\exp(-x_1) + \exp(-x_2) - 1 = 0 \\ &x_1{x_3}^2 - x_6 = 0 \\ &{x_4}^3 - x_1x_3 = 0 \\ &x_5{x_3}^2 + x_1/x_4 - x_6 = 0 \end{aligned} \tag{11}$$

The measurement function (eq 5) is $h = I_6$ (identity matrix of dimension 6), i.e., all the variables are measured. The reconciliation technique is applied with the bounds $\underline{r} = -\bar{r} = 10^{-5}$. Table 1 presents the results with data obtained from the simulated process (system 11). The upper and lower bounds of the measurements are given in rows 2 and 3; their values were reasonably

Table 1. Measurements and Estimations (LMI and LS)

variable	1	2	3	4	5	6
measured lower bound Z	0.67	5.35	4.42	1.54	0.62	18.01
measured upper bound $\bar{z}$	0.91	7.23	5.98	2.08	0.84	24.37
interval center $z_c = (\bar{z} + \underline{z})/2$	0.79	6.29	5.2	1.81	0.73	21.19
interval radius $z_r = (\bar{z} - \underline{z})/2$	0.12	0.94	0.78	0.27	0.11	3.18
estimation $\hat{x}$	0.768	6.215	5.325	1.599	0.751	21.78
correction ratio $100\lvert\hat{x} - z_c\rvert/z_c$	2.4	1.2	2.4	11.7	2.6	2.8
weight $w = z_r/z_c$	0.15	0.15	0.15	0.15	0.15	0.15
LS estimation	0.742	6.468	5.73	1.620	0.728	24.39
correction ratio	5.3	2.3	2.5	12.5	0.4	0.1

Table 2. Parameter Values

parameter	value	parameter	value
V	7.08 m^3	c_p	1.815×10^5 J/kg·°C
V_c	1.82 m^3	c_{pc}	4184 J/kg·°C
A	5.40 m^2	ρ	19.2 kg/m^3
U	3550 J/s·m^2·°C	ρ_c	1000 kg/m^3
K	273.16 °C	k_0	0.0744 m^3/s·kgmol
R	8314.39 J/kgmol	E_0	818 000 J/kgmol
H	-9.86×10^7 J/kgmol		

Table 3. Measurements and Estimations

	true value	measured lower bound	estimation	measured upper bound	correction ratio
F_c	0.0073	0.007	0.007 23	0.008	3.6
F	0.0075	0.007	0.009 92	0.012	−4.5
C_i	2.88	2.15	2.5744	3.0	0
C	1.133	1.1	1.1645	1.25	0.9
T_i	66.075	60	63.7390	70	1.9
T_{ci}	27.00	20	29.3692	30	−17.4
T_c	50.50	45	45.6308	55	8.7
T	88.00	86	89.4017	92	−0.5

Table 4. Balance Residuals before and after Reconciliation

	balance residuals with interval center		
	1	2	3
before equilibration	−0.0001	−0.0062	−0.0048
after equilibration	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$

Table 5. Contributions of the Different Terms of the Second Residual Equation

term	contribution
$\frac{F}{V}(T_i - T)$	−0.0360
$\frac{H}{\rho c_p} k C^2$	−0.0559
$\frac{UA}{V\rho c_p}(T - T_i)$	0.0199

fixed by the user according to available knowledge on the precision of the measurements. A better perception of the measurements can be obtained by considering the center (row 4) and the radius of the measurement interval (row 5). These values give an idea of the mean measurement values and their dispersion.

The estimations are given in the sixth row. The reconciliation was solved by using the iterative scheme of section 3 with initial estimations equal to the centers of the measurement intervals. Only 12 iterations were required to satisfy all of the constraints. The seventh line of Table 1 gives the correction ratio obtained by dividing the magnitude of the correction (difference between the center of the measurement interval and the estimate) by the center of the measurement interval; the magnitudes of these ratios are reasonably small, and therefore, roughly speaking, there are no abnormal data. In other situations, a gross ratio magnitude can be used to detect and localize gross errors or bias in the raw data.[7] All of the mass balances (eq 11) computed with the reconciled values are close to zero, with a precision of 3×10^{-6}, which perfectly agrees with the selected bounds $\underline{r}$ and $\bar{r}$.

The second part of the table gives the results obtained when applying a classical nonlinear least-squares procedure. For that solution approach, we considered as measurements the values obtained by taking the centers of the measurement intervals; for sake of simplicity, the weighting factors w were chosen proportional to the normalized dispersions z_r/z_c (row 5 of the first part of Table 1). The two estimates LMI and LS are in the same vicinity, and the corrective ratios have comparable magnitudes; however, it would be hazardous to go further in this comparison because of the very different concepts of the approaches.

Second Example. This test problem consists of a continuous stirred-tank reactor (CSTR) with a second-order exothermic reaction and heat removal by a coil or jacket. A mathematical model of the CSTR has been developed according to the principle of mass and energy conservation.[30] This example has been used by many authors in the technical literature and constitutes a kind of benchmark for parameter estimation, state estimation, diagnosis, and control. The feed is characterized by the rate F (m^3/s), the concentration of the reactant C_i (kgmol/m^3), and the temperature T_i (°C). The output is characterized by analogous quantities F, C, and T. For the coolant, F_c (m^3/s) and T_{ci} (°C) denote the rate and the temperature, whereas for the output, the same quantities are noted F_c and T_c. Here, we consider the system only under steady-state conditions

$$
\begin{aligned}
0 &= \frac{F}{V}(C_i - C) - kC^2 \\
0 &= \frac{F}{V}(T_i - T) - \frac{H}{\rho c_p} kC^2 - \frac{UA}{V\rho c_p}(T - T_c) \\
0 &= \frac{UA}{V_c \rho_c c_{pc}}(T - T_c) - \frac{F_c}{V_c}(T_c - T_{ci}) \\
k &= k_0 \exp\left[-\frac{E_0}{R(K + T)}\right]
\end{aligned}
\qquad (12)
$$

Table 2 lists the values of the parameters used in the model.

Columns 3 and 5, respectively, of Table 3 contain the lower and upper bounds for the measurements, which were used for the computation of the estimations (column 4). As the process has been simulated, the true values are available and are indicated in column 2. All of the constraints concerning the magnitudes of the corrections are fulfilled, and the three balance equations are closed to zero with a precision of 10^{-5} (Table 4). It is somewhat difficult to analyze the inconstancy and/or the consistency of the raw data through the balance residuals. Indeed, we must bear in mind the magnitudes of the different terms allowing for the calculations of these residuals. It is important to note that residuals are computed as algebraic sums of several quantities; if one of these quantities is negligible with respect to the others, then a bad value for this quantity will have a very small effect on the residual.

For example, we observe that the second mass balance equation in problem 12 is the algebraic sum of three quantities; Table 5 indicates that these three quantities have comparable magnitudes, and consequently, no par-

Table 6. Sensitivity to Bound Modifications

	measured lower bound	estimation	measured upper bound	measured lower bound	estimation	measured upper bound
F_c	0.007	0.007 02	0.008	0.007	0.007 28	0.008
F	0.007	0.009 25	0.012	0.007	0.009 78	0.012
C_i	2.15	2.7473	3.0	2.15	2.6342	3.0
C	1.1	1.1843	1.25	1.1	1.1696	1.25
T_i	63	61.5564	67	60	64.9672	70
T_ci	23	26.9492	27	20	29.1613	30
T_c	45	45.0512	55	45	45.8386	55
T	86	89.2858	92	40	91.4940	150

Table 7. Sensitivity to the Starting Point

	lower bound	interval center	upper bound
F_c	0.007 23	0.007 23	0.007 23
F	0.009 92	0.009 92	0.009 92
C_i	2.5744	2.5745	2.5745
C	1.1646	1.1646	1.1646
T_i	63.7386	63.7377	63.7383
T_ci	29.3701	29.3687	29.3676
T_c	45.6299	45.6313	45.6324
T	89.401 92	89.4019	89.4017

ticular problem of sensitivity occurs in this example. Moreover, the magnitudes of the different terms are significatively greater than the magnitude of the residual (−0.0062, Table 4); thus, the residual can be considered small, suggesting the absence of gross errors in the data.

For analysis of the results, Table 6 exhibits the influence of the modification of some bounds on the reconciliation. The same example is used for two simulations. The left part of Table 6 gives the results for the case where the interval on the temperatures (T_i, T_{ci}) have been reduced by 3 °C, whereas the right part of the table shows a very large increase of the temperature interval T (one can also consider that a large interval might represent the absence of any measurement).

A comparison of the estimations (Tables 5 and 6) shows (for this example) that the sensitivity to the length of the interval is low; in particular, the suppression of a measurement by increasing the corresponding interval, with the restriction that the system remains observable, does not provide any estimation problem.

It is clear, however, that, from a general point of view, the characteristics of an interval (center and length) might influence the estimation. If, for a particular measurement, the corresponding interval is completely separate from the true data, the LMI system becomes inconsistent, i.e., no solution exists. For example, taking the measurement interval $T = [100\ 120]$ does not allow for the existence of a solution. This remark can be applied intensively if it is desired to detect and isolate gross errors affecting the measurements.

Another point to analyze is the convergence of the LMI solution. It is well-known that results obtained by LMI solvers are dependent on the starting point. For the given example, but only for this example, the estimations are weakly influenced by this starting point. The estimations collected in Table 7 were obtained with three different starting points. The first column lists the names of the estimated variables. The second column gives the estimated variables when the procedure is initialized with the lower bound of the measurement interval (see Table 3, third column). The two last columns report the estimations obtained when the initialization uses the center and the upper bound, respectively, of the measurement interval. Thus, it can be concluded that, for this example and also for many others, the state estimation is not very sensitive to the choice of the starting point.

Discussion

As explained in the Introduction, the LMI approach provides a guarantee that the state estimation perfectly satisfies the model constraints and the measurement intervals. Indeed, the method can furnish all admissible solutions satisfying these constraints; however, the numerical resolution of the LMI is often reduced in giving one solution, for example, the solution corresponding to the interior point (see the Appendix). It would be more interesting to enumerate all admissible solutions, those satisfying all of the LMI. This is a tedious task, and the obtained result would be not very convenient for the user.

Let us illustrate the latter point with a toy example. Consider the reconciliation problem

$$x_1 - x_2 - x_3 = 0$$
$$x_1 \in [12\ 14]$$
$$x_2 \in [4\ 6]$$
$$x_3 \in [7\ 9]$$

According to the LMI reconciliation approach, the state estimates are defined as the set of constraints

$$\begin{aligned} &-10^{-5} \le \hat{x}_1 - \hat{x}_2 - \hat{x}_3 \le 10^{-5} \\ &\hat{x}_1 \in [12\ 14] \\ &\hat{x}_2 \in [4\ 6] \\ &\hat{x}_3 \in [7\ 9] \end{aligned} \tag{13}$$

Direct application of the interior approach gives the following (and trivial) results

$$\hat{x}_1 = 13, \qquad \hat{x}_2 = 5, \qquad \hat{x}_3 = 8 \tag{14}$$

In Figure 1a, the solution is given in the plane (x_2, x_3). We have indicated the domain in which the solution lies (defined by constraints 13 and corresponding to the intersection of three strips) and the LMI interior point (eq 14). On one hand, it is clear that the interior point does not reflect the set of admissible solutions; on the other hand, the whole set of admissible solutions expressed by constraints 13 is not very easy to use. A compromise would be to approximate the whole set of solutions (constraints 13) by independent constraints. The reader will verify that the following solution satisfies all of the constaints and is suitable for use

$$\hat{x}_2 \in [4.5\ 5.5]$$
$$\hat{x}_3 \in [7.5\ 8.5]$$

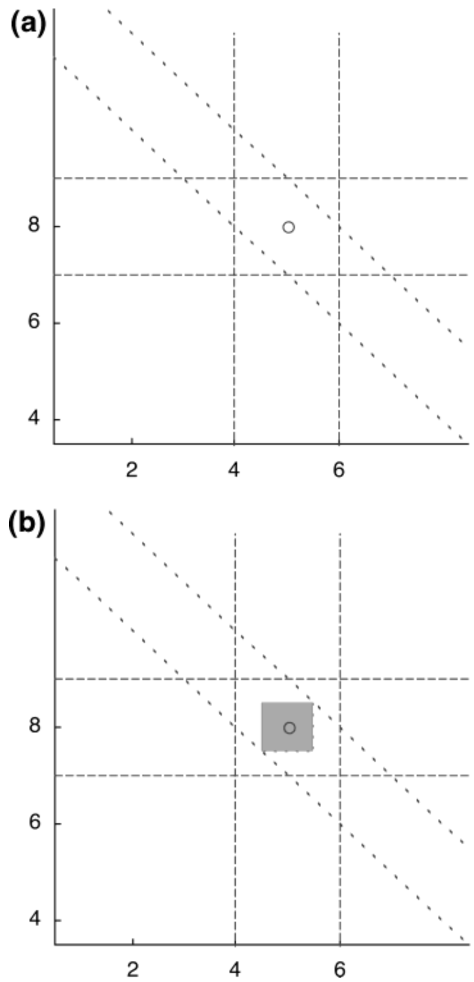

Figure 1. (a) Constraints and interior point solution. (b) Constraints and set of independent solutions.

In Figure 1b, this set of solutions is thus defined by a box (in gray color) in the plane (x_2, x_3). For each solution inside this box, $\hat{x}_1$ can be easily defined by direct substitution $\hat{x}_1 = \hat{x}_2 + \hat{x}_3$.

Summarizing, the point we have discussed with respect to this toy example highlights the fact that the reconciliation problem, when analyzed with an LMI approach, can produce a set of admissible solutions. Moreover, these solutions can be described under a convenient form (independent intervals). The previous discussion might be the starting point for a more elaborate contemplation of the problem of data reconciliation.

5. Conclusion

The technique described herein represents an interesting alternative to the classical technique of data reconciliation using the principle of the maximization of the likelihood function based on the distribution of the measurement errors. It requires very few theoretical hypotheses for implementation and is essentially based on semiempirical knowledge about the variable plausible confidence domains, as well as the definition of a satisfaction threshold for the different balances that depends on the confidence that is placed in them.

We have shown that the classical problems of material balance equilibration can be solved: total balances, partial balances, and balances with missing measurements. The total balance generally refers to the linear case (section 2) and partial balances to the nonlinear case (section 3); to take into account the fact that some measurements are missing, we have considered in the given example that the measurement interval becomes large. Various extensions can be envisaged. The first concerns the development of a reconciliation method that simultaneously uses the precise knowledge (structurally exact balances equations), the imprecise knowledge (balances expressed by inequality constraints), the distribution functions of the errors (when they are available), and the inequality constraints on the correction rate (when the probability distribution functions are unknown). The second extension concerns the integration of fuzzy models or constraints in the form of propositions (for instance, the flow rate of a given stream of the process is "large") or in the form of rules composed of premises and consequences (for instance, if the flow rate of a given stream of the process is "small", then the concentration of the corresponding flow is "high"). That would allow all of the available knowledge on a process to be used with associated weights.

Finally, it would be interesting to examine the case of dynamic systems. A priori, by making an abstraction of the problems associated with the calculation time, the technique proposed here can be applied to such systems because it suffices to express at any given moment the set of constraints (on the rates of correction and on the degree to which the balances must be satisfied) and to find the solution that verifies them. Studies of this extension are currently under way.

Nomenclature

x^* = state variable (true value), dimension n
$\hat{x}$ = state variable (estimation), dimension n
z = state measurement, dimension p ($p \leq n$)
ϵ = error measurement, dimension n
M = incidence matrix, dimension $m \times n$
C = measurement matrix, dimension $p \times n$
$\underline{r}$ = lower bound for the model, dimension m
$\bar{r}$ = upper bound for the model, dimension m
r = common bound for the model, dimension m
$\underline{e}$ = lower bound for the measurement error, dimension p
$\bar{e}$ = upper bound for the measurement error, dimension p
e = common bound for the measurement error, dimension p
f = model equation, dimension m
h = measurement equation, dimension p
I_q = identity matrix, dimension $q \times q$
F = gradient of f, dimension $m \times n$
H = gradient of h, dimension $p \times n$

Appendix: Linear Matrix Inequality

A linear matrix inequality takes the form

$$F(x) = F_0 + \sum_{i=1}^{m} x_i F_i > 0 \qquad \text{(A1)}$$

where $x \in \mathcal{R}^m$, $F_i \in \mathcal{R}^{n\times n}$, $\mathcal{R}^m$ is the set of all real vectors of length m, and $\mathcal{R}^{n\times n}$ is the set of all real $n \times n$ matrices. The symmetric matrices $F_i = F_i^{\mathrm{T}}$, $i = 0, ..., m$, are fixed, and x is the variable. The matrix $F(x)$ is an affine function of the elements of x and is a positive definite matrix, that is, $z^{\mathrm{T}}F(x)z > 0$, $\forall\, z \neq 0$, $z \in \mathcal{R}^n$.

LMI A1 is equivalent to n polynomial inequalities.

Solving LMIs means determining whether the problem is feasible and, if it is, computing a feasible point. Many algorithms for solving LMI problems,[29] including the ellipsoid algorithm are available. Although this might not be the most efficient algorithm, it is very simple to apply.

The idea is as follows: First, we start with an ellipsoid $E^{(0)}$ that is guaranteed to contain the optimal point. Then, we compute a cutting plane (that passes through the center $x^{(0)}$ of $E^{(0)}$) to better localize the optimal point.

To determine this cutting plane, we note that, if $x^{(0)}$ satisfies LMI A1, then there exists a nonzero u such that

$$u^{\mathrm{T}}F(x^{(0)})u > 0 \tag{A2}$$

Define g by

$$g_i = -u^{\mathrm{T}}F_iu \tag{A3}$$

for $i = 1, ..., m$. Then, for any z satisfying

$$g^{\mathrm{T}}(z - x^{(0)}) < 0 \tag{A4}$$

we have

$$\begin{aligned} u^{\mathrm{T}}F(z)u &= u^{\mathrm{T}}F(x^{(0)})u + u^{\mathrm{T}}[F(z) - F(x^{(0)})]u \\ &= u^{\mathrm{T}}F(x^{(0)})u + u^{\mathrm{T}}[\sum_{i=1}^{m}(z_i - x_i^{(0)})F_i]u \end{aligned}$$

From eq A3, one can deduce

$$u^{\mathrm{T}}F(z)u = u^{\mathrm{T}}F(x^{(0)})u + \sum_{i=1}^{m}(z_i - x_i^{(0)})g_i \tag{A5}$$

From eqs A2 and A4, we deduce that $u^{\mathrm{T}}F(z)u > 0$. It follows that every feasible point lies in the half space $\{z|g^{\mathrm{T}}(z - x^{(0)}) < 0\}$, i.e., the vector g defines a cutting plane for the LMI problem at the point x.

We then know that the sliced half-ellipsoid

$$E^{(0)} \cap \{z|g^{\mathrm{T}}(z - x^{(0)}) < 0\}$$

contains the optimal point. We compute the ellipsoid $E^{(1)}$ of minimum volume that contains this sliced half-ellipsoid; $E^{(1)}$ is guaranteed to contain the optimal point. The process is then repeated.

Literature Cited

(1) Ripps, D. L. Adjustment of experimental data. *Chem. Eng. Prog.* **1962**, *58* (10), 120.

(2) Vaclavek, V. Studies on system engineering III. Optimal choice of balance measurements in complicated chemical systems. *Chem. Eng. Sci.* **1969**, 24, p 947.

(3) Smith, H. W.; Ichiyen, N. Computers adjustment of metallurgical balances. *Can. Inst. Min. Metal. Bull.* **1973**, 66, 97−100.

(4) Crowe, C. M.; Garcia Campos, Y. A.; Hrymak, A. Reconciliation of process flow rates by matrix projection. Part 2: The nonlinear case. *AIChE J.* **1983**, *29* (6), 881−888.

(5) Sood, M. K.; Reklaitis, G. W. Solution of material balances for flowsheets modeled with elementary modules: The constrained and the unconstrained case. *AIChE J.* **1979**, *25* (2), 209−229.

(6) Crowe, C. M. Reconciliation identification of gross error in linear data reconciliation. *AIChE J.* **1986**, *32* (4), 616−623.

(7) Narasimhan, S.; Mah, R. S. H. Treatment of general steady-state process models in gross error identification. *Comput. Chem. Eng.* **1989**, 13, 851−853.

(8) Ragot, J.; Darouach, M.; Maquin, D.; Bloch, G. Validation de données et diagnostic. In *Traité des Nouvelles Technologies*; Série Diagnostic et Maintenance; Hermès: Paris, 1990.

(9) Kretsovalis, A.; Mah, R. S. H. Observability and redundancy classification in generalized process networks I. Theorems. *Comput. Chem. Eng.* **1988**, *12* (7), 671−687. Kretsovalis, A.; Mah, R. S. H. Observability and redundancy classification in generalized process networks II. Algorithms. *Comput. Chem. Eng.* **1988**, *12* (7), 689−703.

(10) Crowe, C. M. Observability and redundancy of process data for steady-state reconciliation. *Chem. Eng. Sci.* **1989**, *44* (12), 2909−2917.

(11) Maquin, D.; Darouach, M.; Fayolle, J.; Ragot, J. Localization of sensors in large scale industrial systems. In *Applied Modelling and Simulation of Technological Systems*; Elsevier: New York, 1987: pp 33−39.

(12) Maquin, D.; Luong, M.; Ragot, J. About the design of measurement systems and fault accommodation. *Eng. Simul.* (special issue on Engineering Diagnostics) **1996**, *13*, 1009−1024.

(13) Turbatte, H. C.; Maquin, D.; Cordier, B.; Huynh, C. T. Analytical redundancy and reliability of measurement system. Presented at the IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes, Baden-Baden, Germany, Sep 10−13, 1991.

(14) Kelly, J. D. A regularization approach to the reconciliation of constrained data set. *Comput. Chem. Eng.* **1998**, *22* (12), 1771−1788.

(15) Crowe, C. M. Data reconciliation−Progress and challenges. *J. Process Control* **1996**, *6* (2/3), 89−98.

(16) Crowe, C. M. Formulation of linear data reconciliation using information theory. *Chem. Eng. Sci.* **1996**, *51* (12), 3359−3366.

(17) Ragot, J.; Maquin, M.; Adrot, O. LMI approach for data reconciliation. Presented at the 38th Conference of Metallurgists, Symposium Optimization and Control in Minerals, Metals and Materials Processing, Quebec, Canada, Aug 22−26, 1999.

(18) Romagnoli, J. A.; Sanchez, M. C. *Data Processing and Reconciliation for Chemical Process Operations*; Academic Press: New York, 2000.

(19) Bagajewicz, M. *Process Plant Instrumentation. Design and Upgrade*; Technomic Publishing Company, CRC Press: Boca Raton, FL, 2000.

(20) Narasimhan, S.; Jordache, C. *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data*; Gulf Publishing Company: Houston, TX, 2000.

(21) Mandel, D.; Abdollahzadeh, A.; Maquin, D.; Ragot, J. Data reconciliation by inequality balance equilibration. *Int. J. Min. Process.* **1998**, 53, 157−169.

(22) Maquin, D.; Adrot, O.; Ragot, J. Data reconciliation with uncertain models. *ISA Trans.* **2000**, 39, 35−45.

(23) Himmelblau, D. M. Material balance rectification via interval arithmetic. *Ind. Chem. Eng. Symp. Ser.* **1985**, *92*, 121−133.

(24) Kyriakopoulos, D. J.; Kalitventzeff, B. Validation of measurement data using an interior point SQP. *Comput. Chem. Eng.* **1996**, *20* (suppl.), S563−S568.

(25) Harikumar, P.; Narasimhan, S. A method to incorporate bounds in data reconciliation and gross error detection. Part 2: Gross error detection strategies. *Comput. Chem. Eng.* **1993**, *17* (11), 1121−1128.

(26) Dovi, V. G.; Del Borghi, A. Reconciliation of process flow rates when meausments are subject to detection limits: the bilinear case. *Ind. Eng. Chem. Res.* **1989**, *38* (7), 2861−2866.

(27) Soderstrom, A.; Himmelblau, D. M.; Edgar, T. F. A mixed integer optimization approach for simultaneous data reconciliation and identification of measurement bias. *Control Eng. Pract.* **2001**, 9, 869−876.

(28) Narasimhan, S.; Harikumar, P. A method to incorporate bounds in data reconciliation and gross error detection. Part 1: The bounded data reconciliation problem. *Comput. Chem. Eng.* **1993**, *17* (11), 1115−1120.

(29) Boyd, S. P.; El Ghaoui, L.; Feron, E.; Balakrishnan V. *Linear Matrix Inequalities in System and Control Theory*; SIAM: Philadelphia, PA, 1994.

(30) Smith, C. A.; Corripio, A. B. *Principles and Practice of Automatic Process Control*; John Wiley and Sons: New York, 1985.

(31) Mah, R. S. H. *Chemical Process Structures and Information Flow*; Butterworth Publishers: Stoneham, MA, 1990.

Received for review March 21, 2002
Revised manuscript received November 24, 2003
Accepted December 22, 2003

IE020213S