CHAPTER 3

DATA RECONCILIATION

3.1 Introduction

Process measurements are taken in an industrial plants for the purpose of evaluating process control or process performance. However, not all variables needed are generally measured, because of technical unfeasibility or cost. Furthermore, the measurements often contain random and possibly gross errors as a result of miscalibration or failure of the measuring instruments. Also, the data dose not obey the laws of conservation. The approach that used to reduce the error of the measured data called "data reconciliation" is presented in this chapter.

3.2 Data reconciliation

Process data is the foundation upon which process knowledge, evaluation of process performance and all control system are based. These functions include production planning, operation scheduling, waste treatment scheduling as well as the more familias process control. From above information, mean a correct process data is very important but infact the measured process data inherently contain inaccurate information since the measurements are obtained with imperfect instruments. The relationship between a measured value and correct value can be represented by :

$$y = Y + e \tag{3.1}$$

Where

y = measured value

Y = correct value

e = measurement value

Most measurement systems in an industrial have the same basis problem. The key features of the problems are :

1. All measurement are subject to errors. These errors corrupt the individual measurements and cause the measured values collectively to be consistent in the sense of discrepancies in energy and material balance closure. The flawed-measured value can contain any of several types of error such as :

1.1. Small random errors

These errors are commonly assumed to be independently, normally distributed with zero mean and expected value of the errors is also zero.

1.2. Systematic biases

These errors occur when measurement devices provide consistently erroneous values, either high or low. In this case, the expected value of the errors is not zero. Biases may arise from sources such as incorrect installation or calibration of the measurement device.

1.3. Gross errors

These errors are usually caused by non-random events. In this case, the measurement bear little or no relation to the true value of the desired property. Gross errors can be subdivided into measurement-related errors such as malfunctioning sensors and process-related errors such as process leaks.

2. Not all process variables are measured for reasons of cost, inconvenience or technical unfeasibility.

3. Redundancy, in general, means providing two or more process variable measurements, computers, sensors, pumps, valves and so on, that are design for the same function. Redundant measurement means obtaining the same process information with two or more measurements. Redundant measurement can typical into two type.

First, spatially redundant, measurements are spatially redundant if there are more than enough data to completely define the process model by a result of conservation relationship and the interconnectedness of a process network.

Second, temporally redundant, measurements are temporally redundant if the process conditions were truly at a steady state, more measurements are available than need.

Dynamic models comprised of algebraic and differential equations provide both spatial and temporal redundancy.

A simplified view of techniques used to process measurement data can be divided into three basic steps as shown in Figure 3.1.

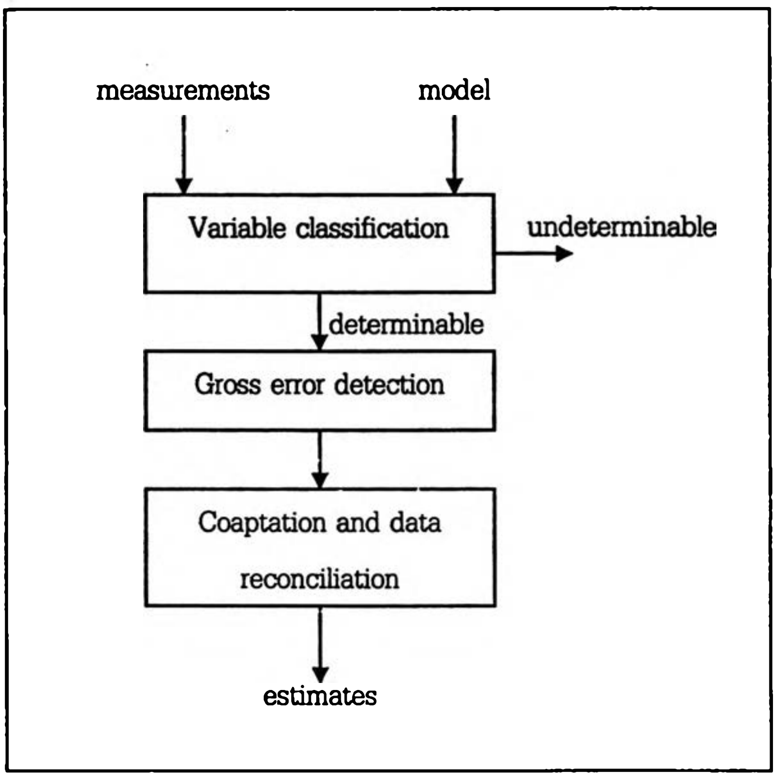

Figure 3.1 Steps for processing measurement data

1. **Variable classification, the first step** that is involved determining which variables are observable or unobservable and redundant or underdetermined. Variables which are undeterminable are not available for improvement.
2. **Gross error detection, the second step** that is the step for identifying and removing all gross errors. Gross error detection must be performed prior to data reconciliation step because if a measurement containing a gross error

were allowed into the reconciliation scheme, the resulting variable estimates would contain significant errors, with the entering gross error accounted for in some or perhaps all of the estimates.

3. **Data reconciliation and Coaptation, the third or final step** that is used to improve process knowledge from process measurement data. This step obtains consistent estimates which satisfy all of the specified model equations while staying as close as possible to the actual measurement. Coaptation step estimates all unmeasured but observable variables that can be treated simultaneously with measured data reconciliation.

The first to propose a data reconciliation algorithms in the chemical engineering literature was a algorithm that based on a minimal least squares adjustment of the measurements subject to a number of reconciliation equations. This algorithm did not even model gross or systematic errors (Kuehn and Davidson, 1961). And the next research pointed this out, and developed an algorithm that incorporated a mechanism to detect gross error (Ripps, 1965). Since, many algorithms for gross errors detection had been suggested. Data reconciliation techniques could be divided in six classes (Himmelblav, 1988).

1. Techniques based on the least squares principle (maximum likelihood) represented the largest group. These techniques were also called Gauss-Markov estimates if the weights correspond to the inverse of the variance of each response. Techniques based on the least squares principle were most frequently reported on, most thoroughly tested and best known. In the simple terms, one could say that they minimize a sum of squares of measurement corrections, usually with step to detect systematic errors and replace them by rectification or reconstruction. Attention would be focused on this type of technique become it was the most widely used and studied.
2. Some techniques based on mathematical programming, in a more general way than least squares application as specified in (1), were proposed as well. For instance, the technique that based on the linear sets of

reconciliation equations (Marro et. al., 1981). In some researches suggested the algorithms that based on The Lagrange multipliers (e.g. : Iordache *et. al.*, 1985, and Serth *et. al.*, 1986).

3. Techniques based on Kalman filtering were used, especially for quasi steady state systems (e.g. : Stanley and Mah, 1977, and Stanley, 1982). A quasi steady state variable was defined as a variable that could change only very slowly, or had an occasional sharp transition in between the two steady state values.
4. Techniques based sensitivity coefficients were used, minimizing the sensitivity of the reconciied measurements for potential errors (Vaclavek *et. al.*, 1979).
5. A method of data reconciliation based on data from which gross errors were removed, using an interval analysis (Himmelblav, 1988). The advantage of this method was that no assumption had to be made about the statistical distribution of the random measurement errors. A disadvantage was that interval analysis was not very well known among engineers, the major practitioners of data reconciliation.
6. Some methods were based on an artificial intelligence principles. These methods only pertained to very specific reconciliation methods, and were of no general interest.

It was important to stress that many of the published reconciliation algorithms did not take constraints into account. This could lead to unrealistic results, thus, the constraint data reconciliation were developed in many different approaches.

An extensive analysis of data reconciliation with solution in matrix form for linear constraints, including unmeasured flows and products of measured flows and measured concentrations was presented (Swenker *et. al.*, 1964). The constraint equations might be written as :

$$B_1(x' + a) + B_2(d' + b) + Pv = 0 \qquad (3.2)$$

where, x' and d' were the vector of measured species, a and b were the corresponding vector of adjustments.

A method to handle data reconciliation in a decomposed scheme, with out iteration based on the assumption that every plants could be divided into a number of departments, with measurement vector u_i for the i^{th} department. The departments were connected with streams v and the reconciliation equations within the departments as well as for the interconnecting streams, were linear (Marro *et. al.*, 1981). This data reconciliation problem was formulated by the following equations :

$$\min_{\eta,\varsigma} \left(\sum_{i=1}^{N} \eta_i^T W_i \eta_I + \varsigma^T W_0 \varsigma \right) \tag{3.3}$$

subject to :

$$D_0(v+\varsigma) = 0 \tag{3.4}$$

$$B_i(u_i+\eta_i) + D_i(v+\varsigma) = 0\ , \quad \forall i = 1,\ldots,N \tag{3.5}$$

This technique did not reduce the amount of calculations. It suggested a technique for spreading the computational time and reducing the amount of data that had to be transferred. The quasi steady state data reconciliation was presented in the sense of spreading the computational time (Stanley, 1982).

The other applications of the data reconciliation problem are: A dynamic on-line estimation algorithm for reconciling process variables that involved a recursive solution technique in weighted least squares (Darouach and Zasadyinski, 1991), a method to incorporate bounds in data reconciliation that the bounds on process variables were directly incorporated as constraints (Narasimhan and Haukumar, 1993), a case study of data reconciliation in Kultin ammonia plant (Nugrah et. al., 1993), and an application of data reconciliation to an industrial pyrolysis reactor that developed around simplified mass and energy balances (Weiss et. al., 1996). The data reconciliation algorithm that produced for the acetylene hydrogenation process in this thesis is the one of the applications of the data reconciliation problem.

3.3 Dynamic data reconciliation problem formulation

The general nonlinear dynamic data reconciliation (NDDR) formulation can be written as:

$$\min_{y} \Phi[Y; y; \sigma] \qquad \text{(A)}$$

subject to

$$
\begin{aligned}
f\left[\frac{dy(t)}{dt}; y(t)\right] &= 0 \\
h[y(t)] &= 0 \\
g[y(t)] &\geq 0 \\
y(t)^L \leq y(t) &\leq y(t)^U
\end{aligned}
\qquad (3.6)
$$

Where

ϕ = objective function equation

Y = discrete measurement

y = estimate function

y^L = lower limit of y

y^U = upper limit of y

σ = measurement noise standard deviation

f = differential equation constraints

h = algebraic equality constraints

g = inequality constraints including simple upper and lower bounds

The nonlinear dynamic data reconciliation (NDDR) approach is composed with several steps :

1. Obtain process measurements over $t_c - H\Delta t \leq t \leq t_c$.
2. Estimate *y(t)* that minimize ϕ over $t_c - H\Delta t \leq t \leq t_c$
3. Save only *y(t)* at time t_c .
4. Move to the next calculation time : t_c -----> $t_c + \Delta t = t_{c+1}$

5. Repeat step 1-3 at next time over, $t_{c+1}-H\Delta t \leq t \leq t_{c+1}$
6. Repeat step 4

As shown in Figure 3.2, only data measurements within the horizon will be reconciled during the nonlinear dynamic data reconciliation run.

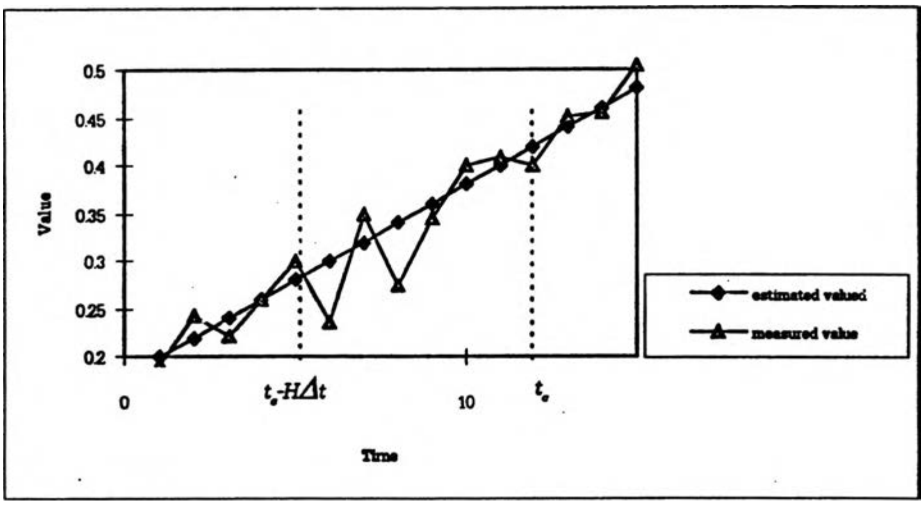

Figure 3.2 History horizon for NDDR

The objective function can be rewritten in the form of the Weighted Least-Squares (WLS) formulation :

$$\phi\left[Y;y;\sigma\right] = \sum_{j=tc-Ht}^{tc} \frac{1}{2}\left[y_j - Y_j\right]^T V^{-1}\left[y_j - Y_j\right] \quad (3.7)$$

or

$$\phi = \frac{1}{2}\left[y - Y\right]^T V^{-1}\left[y - Y\right] \quad (3.8)$$

Where

y_j = the estimated value of estimation equation at time t_j

Y_j = the measured value at time t_j

t_c = the current time

Δt = time step size

H = the history horizon time

V = the variance-covariance matrix

The concept of data reconciliation is to : minimize ϕ or minimize the sum of errors between measured values and reconciled values by using the Weighted Least-Squares Estimation.

$$\phi = \frac{1}{2}[y - Y]^T V^{-1}[y - Y] \tag{3.9}$$

from $(A+B)^T = A^T + B^T$ obtain

$$\frac{\partial\phi}{\partial y} = \frac{1}{2}\left\{2y^T V^{-1} - 2Y^T (V^{-1})^T\right\} = 0 \tag{3.10}$$

$$y^T V^{-1} - Y^T (V^{-1})^T = 0 \tag{3.11}$$

$$y^T = \frac{Y^T (V^{-1})^T}{V^{-1}} \tag{3.12}$$

$$y = \left(\frac{Y^T (V^{-1})^T}{V^{-1}}\right)^T \tag{3.13}$$

For example: If

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \tag{3.14}$$

$$y = \begin{bmatrix} a1Y_1 + a2 \\ b1Y_2 + b2 \end{bmatrix} \tag{3.15}$$

$$V = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \tag{3.16}$$

$$\phi=\frac{1}{2}\begin{bmatrix} a1Y_1+a2-Y_1 \\ b1Y_2+b2-Y_2\end{bmatrix}^T\begin{bmatrix}\sigma_1 & 0 \\ 0 & \sigma_2\end{bmatrix}^{-1}\begin{bmatrix} a1Y_1+a2-Y_1 \\ b1Y_2+b2-Y_2\end{bmatrix} \tag{3.17}$$

$$\phi=\frac{1}{2}\left(\frac{1}{\sigma_1}\left(a1Y_1+a2-Y_1\right)^2+\frac{1}{\sigma_2}\left(b1Y_2+b2-Y_2\right)^2\right) \tag{3.18}$$

Minimize ϕ by adjusting the a1, a2, b1, and b2 parameter. The proper value of a1, a2, b1, and b2 can be found by solving the following equations:

$$\frac{\partial\phi}{\partial a1}=\frac{1}{\sigma_1}\left(2(a1-1)Y_1^2+2a2Y_1\right)=0 \tag{3.19}$$

$$\frac{\partial\phi}{\partial a2}=\frac{1}{\sigma_1}\left(2(a1-1)Y_1+2a2\right)=0 \tag{3.20}$$

$$\frac{\partial\phi}{\partial b1}=\frac{1}{\sigma_2}\left(2(b1-1)Y_2^2+2b2Y_2\right)=0 \tag{3.21}$$

$$\frac{\partial\phi}{\partial b2}=\frac{1}{\sigma_2}\left(2(b1-1)Y_2+2b2\right)=0 \tag{3.22}$$

For the constraint problems, the Lagrange Multiplier method is involved for solving the problems. The procedure of using the Lagrange Multiplier is adding the constraints into the objective equation, thus, ϕ becomes $\phi + w\,h(y(t))$ or

$$\text{new } \phi = L(y,w) = \phi + w\,h(y(t)) \tag{3.23}$$

The proper value of parameter of the estimation equation can be found by solving the following equations:

$$\frac{\partial L}{\partial a1}=\frac{1}{\sigma_1}\left(2(a1-1)Y_1^2+2a2Y_1\right)=0 \tag{3.24}$$

$$\frac{\partial L}{\partial a2} = \frac{1}{\sigma_1}\left(2(a1-1)Y_1 + 2a2\right) \quad = 0 \tag{3.25}$$

$$\frac{\partial L}{\partial b1} = \frac{1}{\sigma_2}\left(2(b1-1)Y_2^2 + 2b2Y_2\right) = 0 \tag{3.26}$$

$$\frac{\partial L}{\partial b2} = \frac{1}{\sigma_2}\left(2(b1-1)Y_2 + 2b2\right) \quad = 0 \tag{3.27}$$

and

$$\frac{\partial L}{\partial w} = h(y(t)) \quad = 0 \tag{3.28}$$

3.4 Summary

The reviews and theories of the data reconciliation approach that are presented in this chapter gives the basic understanding of the data reconciliation approach. The understandind will be used to apply with the acetylene hydrogenation process. The developed dynamic data reconciliation of the acetylene hydrogenation process is presented in chapter next chapter 7. Anyway, the next chapter presents the Dynamic Matrix Control theory.