Module:
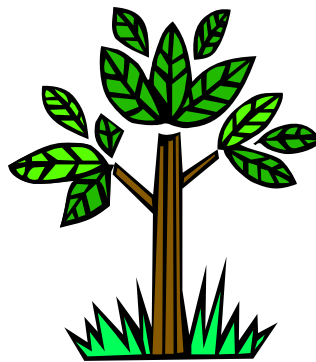
Introduction to Data Reconciliation

Program for North American Mobility in Higher Education

**Introducing Process Integration for
Environmental Control in Engineering Curricula**

Objective of this Module:

This module introduces concepts of data reconciliation techniques that have been widely used in processing industries. Some examples are presented throughout the text, so that an audience can easily understand the data reconciliation algorithms. After completing this module, an audience should be able to solve practical problems encountered in reconciling process data, either by their coded MATLAB programs, or commercial software.

Structure of This Module

This module consists of three tiers:

■ **Tier I – Basic Concepts in Data Reconciliation**

■ **Tier II – Case Studies**

■ **Tier III – Open-Ended Problem**

Each tier is ideally completed in order. Some practical examples and quizzes are presented throughout the module to better gasp the various concepts. For each quiz, there are one or more than one correct answers.

Tier 1:

Basic Concepts in Data Reconciliation

Basic Concepts in Data Reconciliation

Table of Contents

Basic Concepts in Data Reconciliation

Table of Contents (cont'd)

Basic Concepts in Data Reconciliation

Table of Contents (cont'd)

Basic Concepts in Data Reconciliation

Table of Contents (cont'd)

Basic Concepts in Data Reconciliation

Chapter 1:
General Introduction

Basic Concepts in Data Reconciliation

1.1 Process Measurements

Measured process data inevitably contain some inaccurate information, since measurements are obtained with imperfect instruments which have their own accuracy. In addition, signal transmission, power fluctuation, improper instrument installation and miscalibration are other sources of measurement errors.

It is assumed that any observation is composed of a true value plus some error value. This indicates that a measurement can be modeled as:

$$y = x + e \qquad\qquad (1.1)$$

where y is the observed value of the raw measurement, x is the true value of the process variable, and e is the measurement error.

Basic Concepts in Data Reconciliation

1.2 Measurement Error

The error term in Equation (1.1), e, can be divided into two subcomponents, random error and gross error, as shown in Figure 1.1.

$$y = x + e$$

Two Components:

■ Random Error, ε

■ Gross Error, e_s

$$y = x + e_s + \varepsilon$$
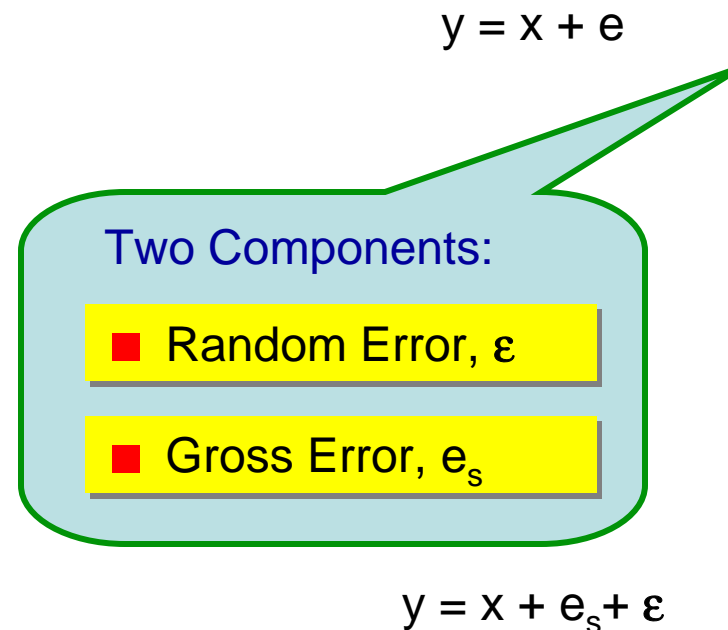
Figure 1.1: Components of measurement errors

Basic Concepts in Data Reconciliation

1.2 Measurement Error

Random error is caused by one or more factors that randomly affect measurement of a variable. It follows a Gaussian distribution.
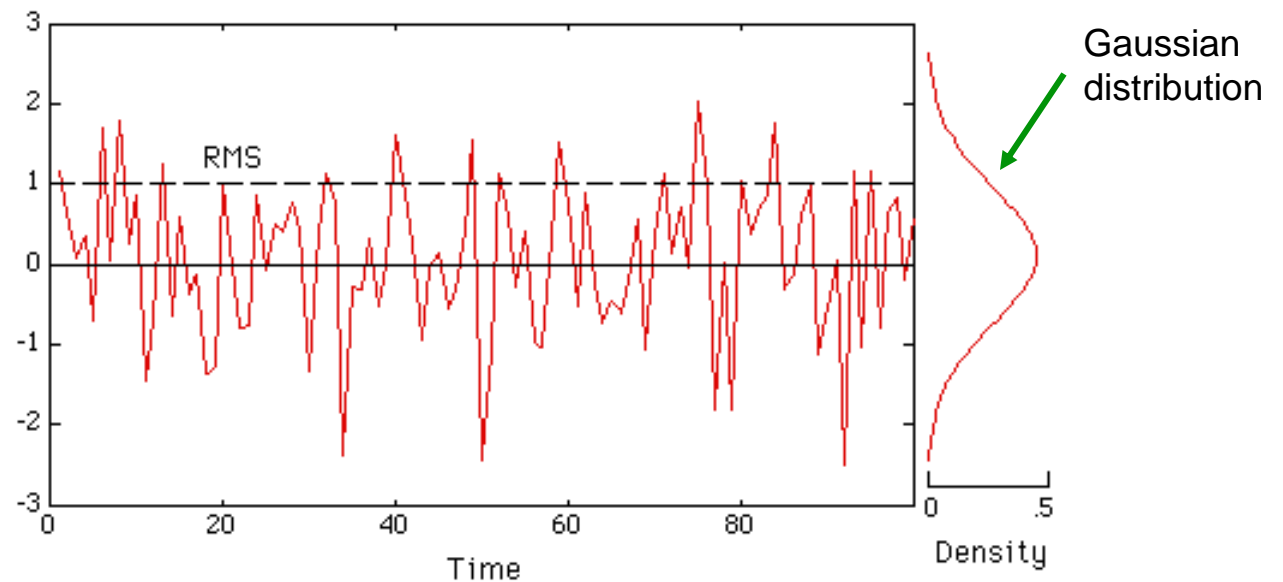
Figure 1.2: Typical measurement errors as Gaussian noise

Basic Concepts in Data Reconciliation

1.2 Measurement Error

The Gaussian noise is normally distributed with a mean value of zero and known variance. The probability density function (PDF) of a measurement with Gaussian noise is described by the formula:

$$P(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y\text{-}\mu)^2}{2\sigma^2}\right) \tag{1.2}$$

where μ is the mean value of the measurements, and σ is the standard deviation.

The important property of random error is that it adds variability to the data, but it does not affect average performance for the group.

Basic Concepts in Data Reconciliation

1.2 Measurement Error

Gross error (as depicted in Figure 1.3) can be caused by:

■ instrument systematic bias that is consistently erroneous, either higher or lower than the true value of the process variable, probably because of instrument miscalibration

■ measurement device failure

■ nonrandom events affecting process, such as process leak.
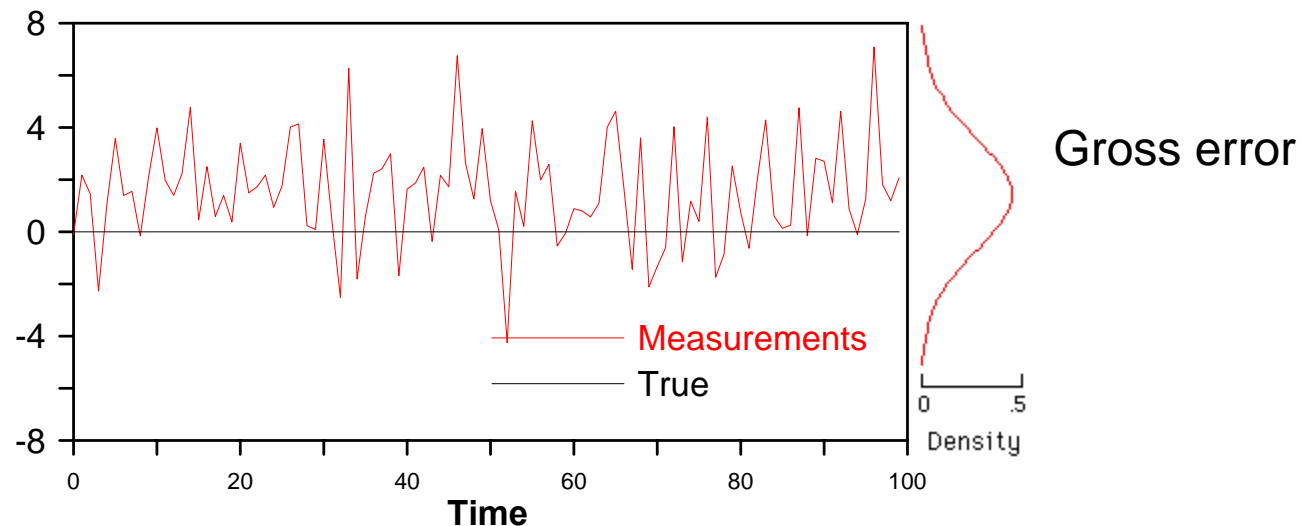
Figure 1.3: Gross error in measurements

Basic Concepts in Data Reconciliation

1.2 Measurement Error

Unlike random errors, gross errors tend to be consistently either positive or negative. Because of this, it is sometimes considered to be a bias in the measurement.

Generally, measurements with gross errors will lead to severely incorrect information about the process, much more so than those with random errors. Gross error detection is an important aspect in validation of process data, and will be discussed further in Chapter 8.

Basic Concepts in Data Reconciliation

1.2 Measurement Error

Errors in measured data can lead to significant deterioration in plant operation. Small random and gross errors deteriorate the performance of control systems, whereas larger gross errors can nullify process optimization. It is important to estimate the true conditions of process states with the information provided by the raw measurements, in order to achieve optimal process monitoring, control, and optimization.
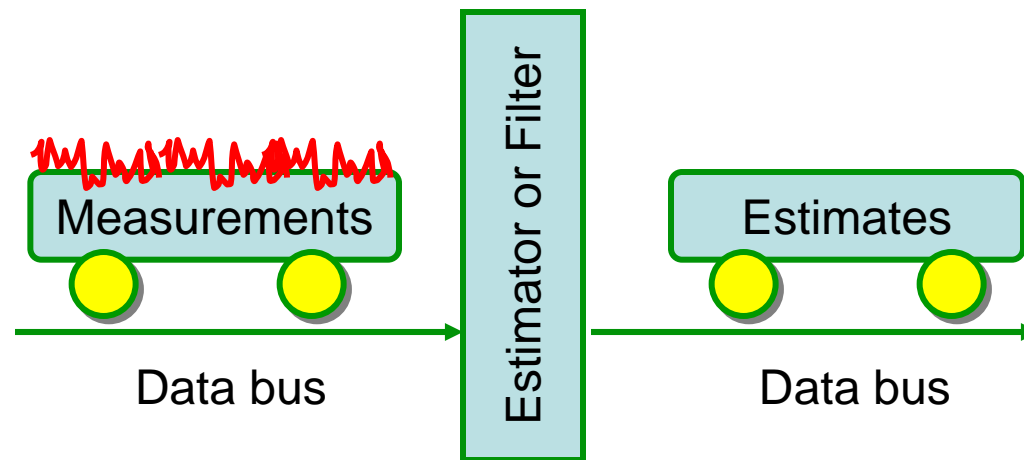
Basic Concepts in Data Reconciliation

1.3 Data reconciliation

The estimation of a process state involves the processing of the raw data and their transformation into reliable information.

For example,
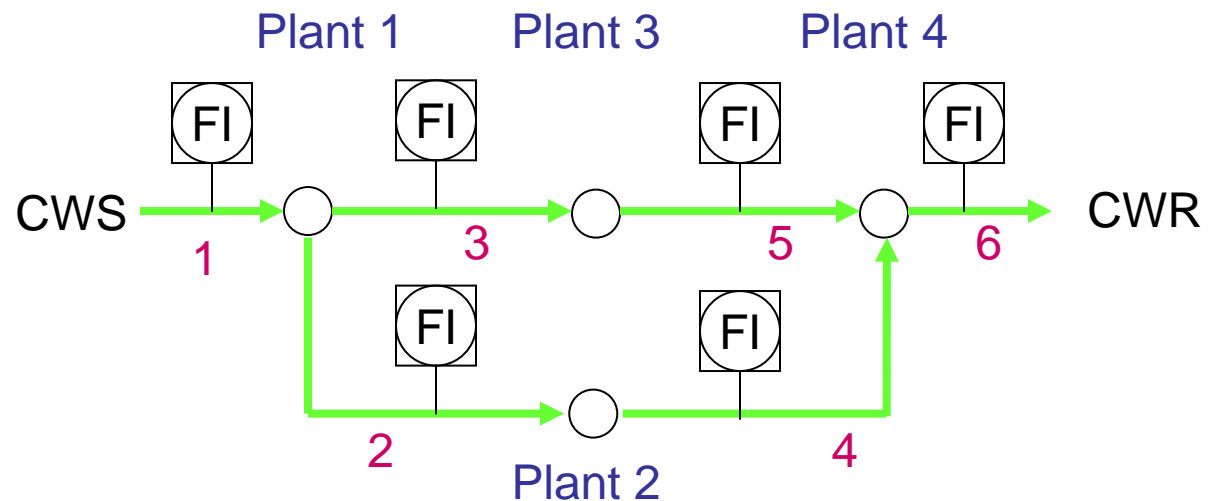
CWS: Cooling Water Supply

CWR: Cooling Water Return

Figure 1.4: A cooling-water circulation network

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

a cooling-water station provides water for four plants as shown in Figure 1.4. All the flow rates for the circulation water are measured in this network. At steady-state, the raw measurements and their standard deviations are listed in Table 1.1.

Figure 1.4

Table 1.1: Flow measurements in cooling water network

Stream No.	Raw Measurement (kt/h)	Standard Deviation, σ (kt/h)
1	110.5	0.82
2	60.8	0.53
3	35.0	0.46
4	68.9	0.71
5	38.6	0.45
6	101.4	1.20

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

If we make mass balances around each plant in the network using the raw measurements, we will find that all the flow measurements contain errors. This is because the true values of the flow rates must satisfy mass balances at steady state. For example, the measurement of stream 1, coming into Plant 1, is 110.5 kt/h. However, the sum of the measured flows for streams 2 and 3 leaving Plant 1 is 60.8 + 35.0 = 95.8 kt/h. Now the question is, how many tons of cooling water does each plant use? For Plant 1, is it 110.5 kt/h or 95.8 kt/h?

The estimation of the true values for the flows in this network can be solved by Date Reconciliation (DR).

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

Data reconciliation is the estimation of process variables based on information contained in the process measurements and models. The process models used in the data reconciliation are usually mass and energy conservation equations.
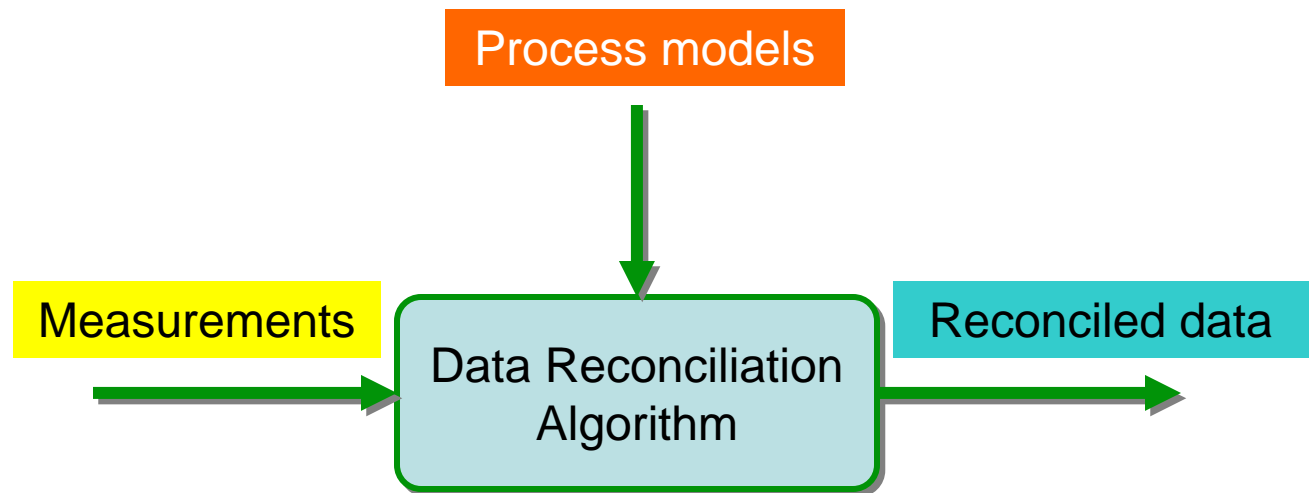
```
                    ┌──────────────────┐
                    │  Process models  │
                    └──────────────────┘
                             │
                             ▼
┌──────────────┐    ┌──────────────────┐    ┌──────────────────┐
│ Measurements │───▶│  Data Reconciliation  │──▶│ Reconciled data │
└──────────────┘    │     Algorithm    │    └──────────────────┘
                    └──────────────────┘
```

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

The DR technique allows the adjustment of the measurements so that the corrected measurements are consistent with the corresponding balances. This information from the reconciled data can be used by the company for different purposes such as:

- Monitoring
- Optimization
- Simulation
- Instrument maintenance

- Management
- Modeling
- Control
- Equipment analysis

This is especially true with the implementation of a Distributed Control System (DCS), as shown in Figure 1.5.

Basic Concepts in Data Reconciliation

1.3 Data reconciliation
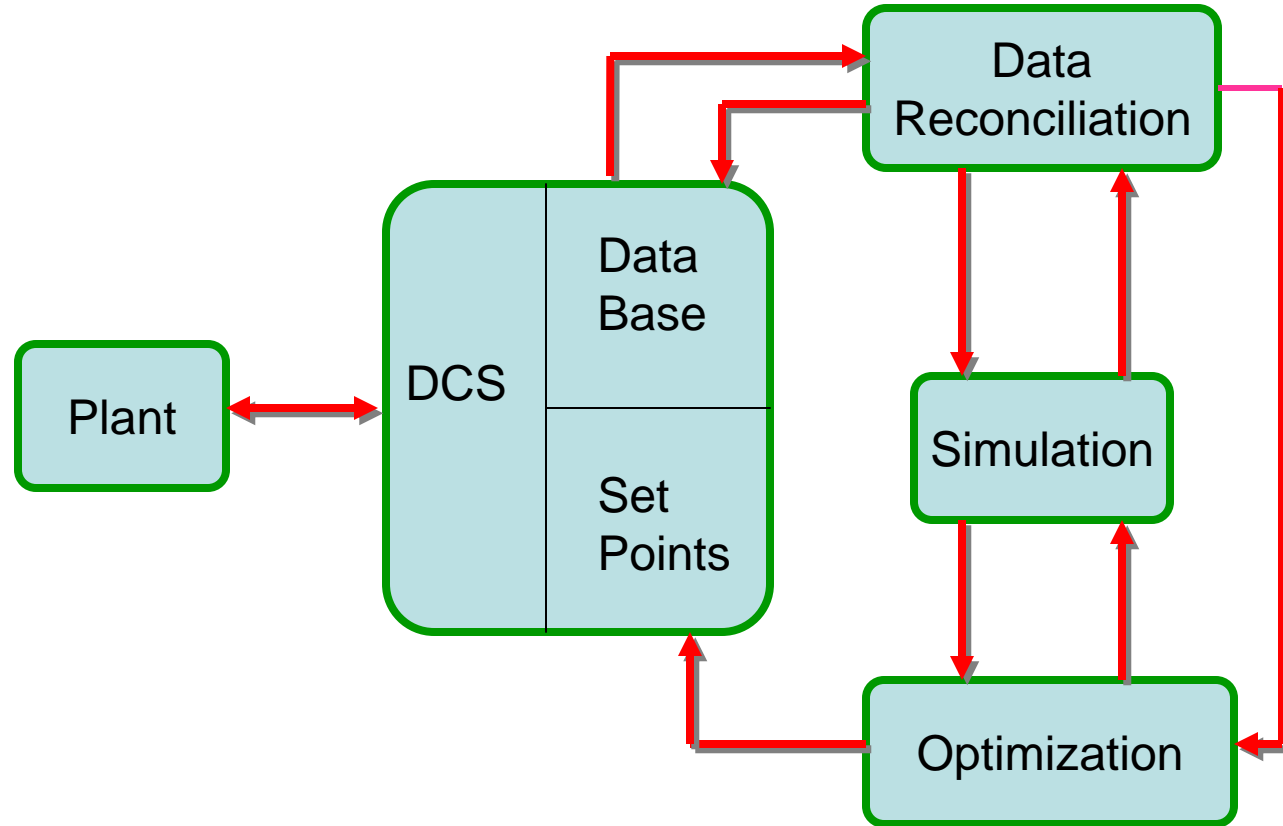
Figure 1.5: Interconnections between data reconciliation, process simulation, and optimization.

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

The interest in applying DR techniques started in the 1980's when plant management realized the benefits of having access to more reliable estimates of process data. Nowadays, data reconciliation techniques have been widely applied to various processing industries, such as:

- Refinery
- Petrochemical
- Metal/Mineral
- Chemical
- Pulp/Paper

Commercial software specializing in data reconciliation is available. A demo-version of one commercial software can be downloaded at: http://www.simsci.com/products/datacon.stm.

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

Research and development during the past 30 years have led to two major types of applications:

■ Mass and heat balance reconciliation. The simplest example is the off-line reconciling of flow rates around process units. The reconciled flow rates satisfy the overall mass balance of the units.

■ Model parameter estimation. Accurate, precise estimates of model parameters are required in order to obtain reliable model predictions for process simulation, design and optimization. One approach to the parameter estimation is to solve the estimation problem simultaneously with the data reconciliation problem. The reconciled model parameters are expected to be more accurate and can be used with greater confidence.

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

In general, the optimal estimates for process variables by DR are solutions to a constrained least-squares or maximum likelihood objective function, where the measurement errors are minimized with process model constraints.

With the assumption of normally distributed measurements, a least-squares objective function is conventionally formulated for the data reconciliation problem. At process steady state, the reconciled data are obtained by:

minimizing $J(\hat{\mathbf{y}}, \hat{\mathbf{z}}) = (\mathbf{y} - \hat{\mathbf{y}})^\top \mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$ (1.3)

subject to $\mathbf{f}(\hat{\mathbf{y}}, \hat{\mathbf{z}}) = \mathbf{0}$

$\quad\quad\quad\quad \mathbf{g}(\hat{\mathbf{y}}, \hat{\mathbf{z}}) \geq \mathbf{0}$

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

where

y is an M×1 vector of raw measurements for M process variables,

ŷ is an M×1 vector of estimates (reconciled values) for the M process variables,

ẑ is an N×1 vector of estimates for unmeasured process variables, **z,**

V is an M ×M covariance matrix of the measurements,

f is a C×1 vector describing the functional form of model equality constraints,

g is a D×1 vector describing the functional form of model inequality constraints which include simple upper and lower bounds.

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

The models employed in DR represent variable relationships of the physical system of the process. The reconciled data takes information from both the measurements and the models. In reconciling steady-state measurements, the model constraints are algebraic equations. On the other hand, when dealing with dynamic processes, dynamic models that are differential equations have to be used.

Based on the type of model constraints, the data reconciliation problem can be divided into several sub-problems as shown in Figure 1.6. Each sub-problem will be discussed respectively in this module.

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

CHAPTER 1

General Introduction
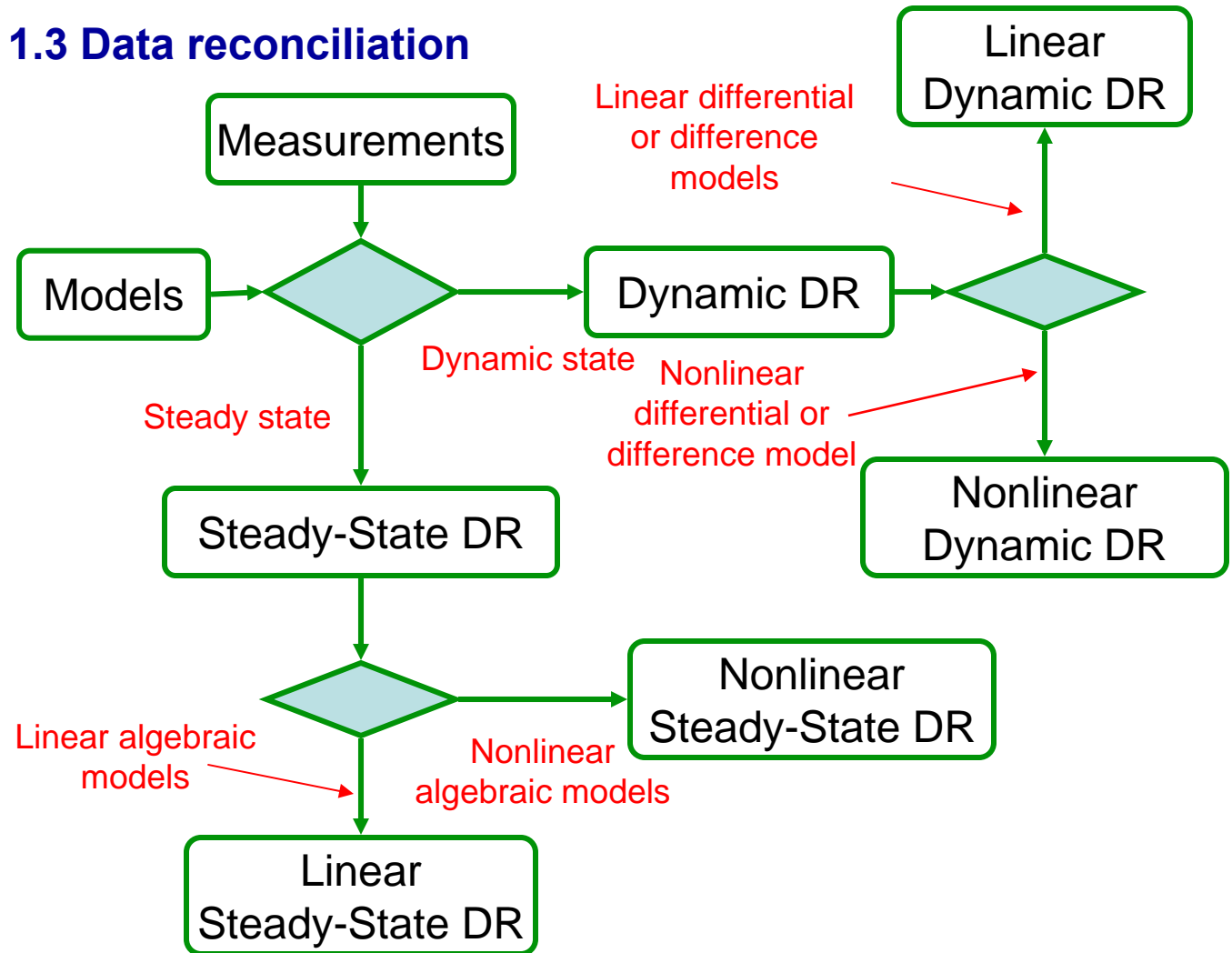
Figure 1.6: Subproblems in data reconciliation

Basic Concepts in Data Reconciliation

1.3 Data reconciliation

The algorithm of the DR formulated by Equation (1.3) indicates that the data reconciliation techniques not only reconcile the raw measurements, but also estimate unmeasured process variables or model parameters, provided that they are observable.
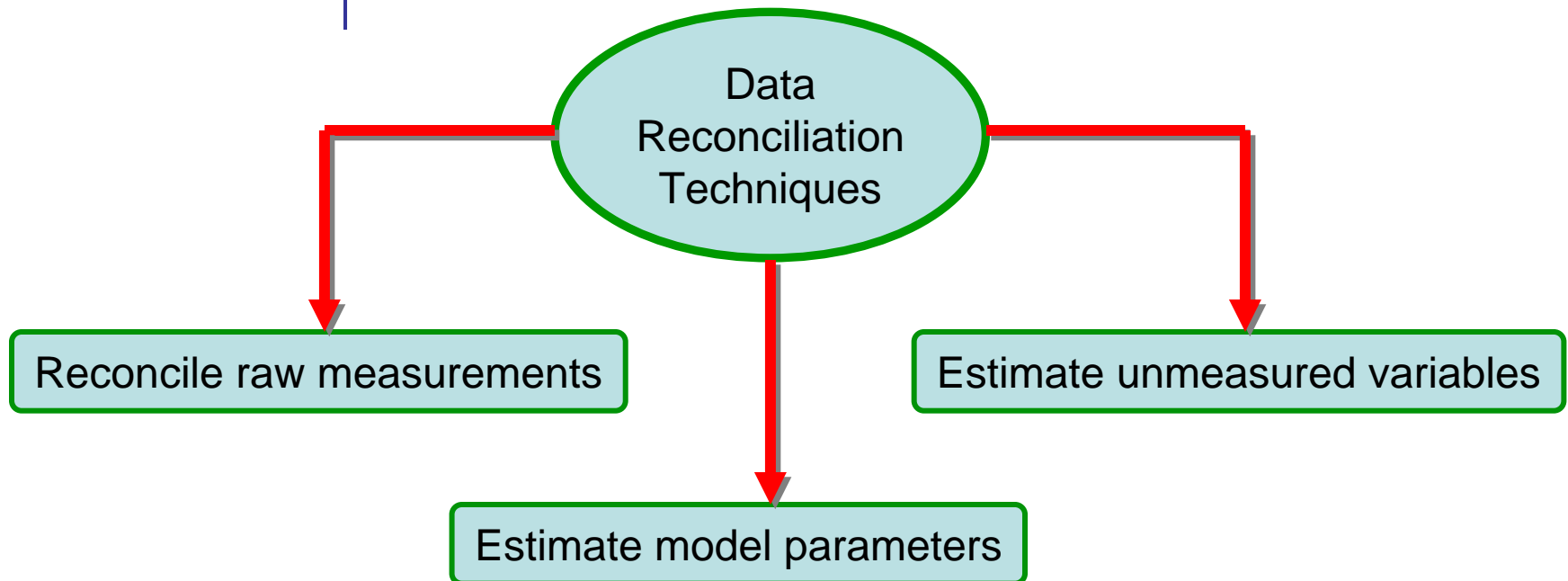
Data Reconciliation Techniques

Reconcile raw measurements

Estimate model parameters

Estimate unmeasured variables

Basic Concepts in Data Reconciliation

1.4 Process Variable Classification

It is also important to clarify some concepts in DR techniques Measured variables are classified as redundant and nonredundant, whereas unmeasured variables are classified as observable and nonobservable. The classification of process variables is shown in Figure 1.7.
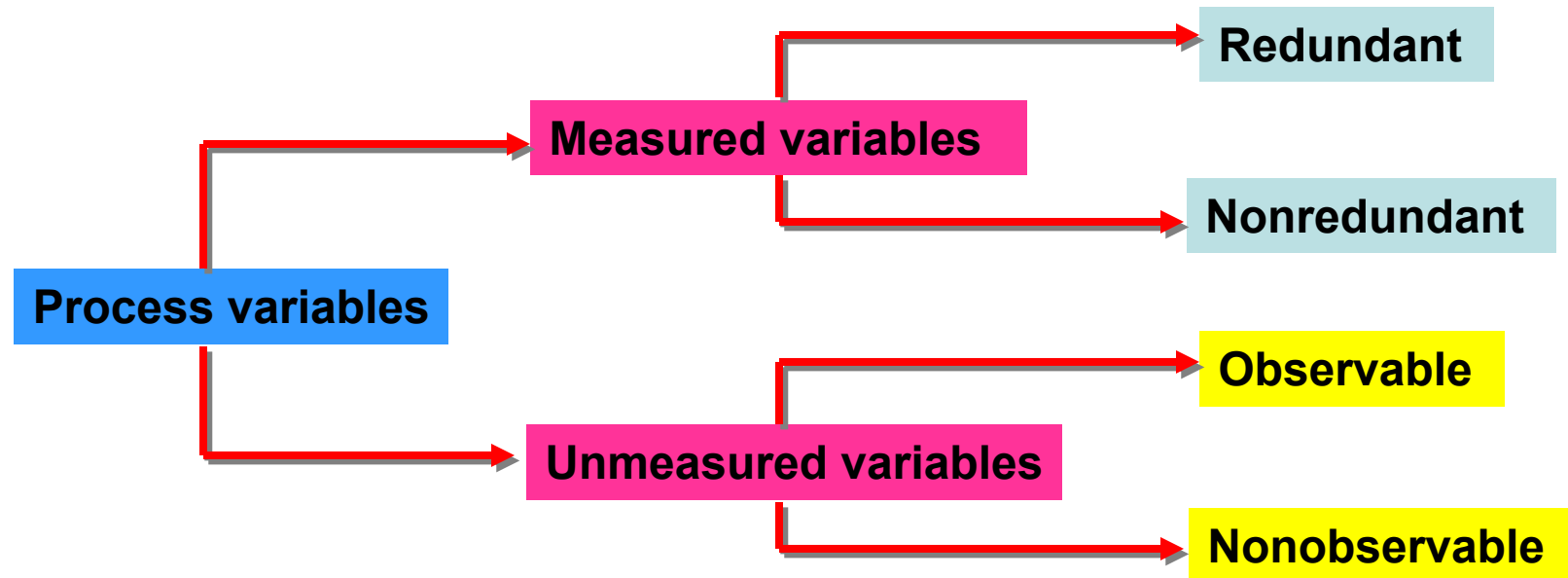
Figure 1.7: Classification of process variables

Basic Concepts in Data Reconciliation

1.4 Process Variable Classification

■ A redundant variable is a measured variable that can be estimated by other measured variables via process models, in addition to its measurement.

■ A nonredundant variable is a measured variable that cannot be estimated other than by its own measurement.

■ An observable variable is an unmeasured variable that can be estimated from measured variables through physical models.

■ A nonobservable variable is a variable for which no information is available.

Basic Concepts in Data Reconciliation

1.4 Process Variable Classification

To demonstrate these concepts, we take the cooling water network as the example:

In Figure 1.4, all six flows are measured, and any one of them can be estimated by mass balances using other measured flows, so they are all redundant variables.

However, if the measurements of flows 2, 4, and 6 were eliminated as shown in Figure 1.8, flow 1 becomes a measured nonredundant variable, but the measurements of flows 3 and 5 are redundant. The unmeasured flows 2, 4, and 6, in this case, are observable, because their values can be estimated by mass balances around the plants, using the measured flows.

Basic Concepts in Data Reconciliation
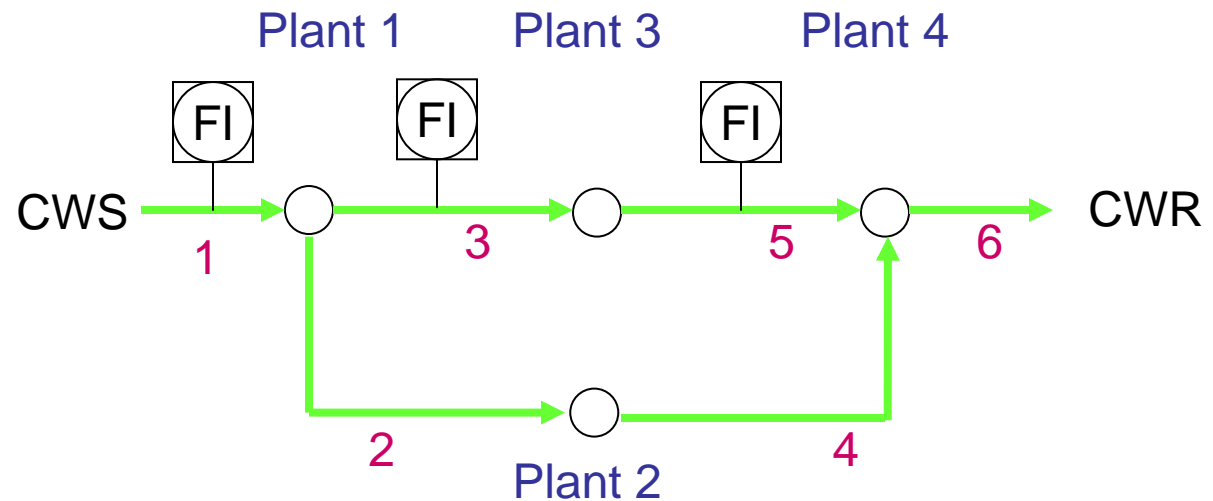
Figure 1.8: Cooling water network with measurements of flows
2, 4, and 6 eliminated.

Basic Concepts in Data Reconciliation

1.5 Redundancy

■ A measurement is spatially redundant if there are more than enough data to completely define the process at any instant in time. Referring to Figure 1.4, all the measurements are spatially redundant. For example, we don't need the value of the measurement for flow stream 1, we can still completely define the process. This is because flow stream 1 can be calculated by other spatial measurements via mass balances.

■ A measurement is temporally redundant if its past measurements can be used to estimate the current state. A typical case for a temporally redundant measurement is that, at the current sampling time, t, the true value of the process variable can be predicted by dynamic models, in addition to the raw measurement.

Basic Concepts in Data Reconciliation

1.6 Quiz

Question 1:

A measurement

(a) may contain random error and/or gross error.

(b) is always perfect.

(c) is a random variable.

(d) is a deterministic variable.

Question 2:

The effects of a systematic measurement bias on the estimation of a process

(a) is more significant than that of random error.

(b) can be negligible compared with that of random error.

(c) is compatible to a random error.

(d) can be eliminated provided that it is detected.

Basic Concepts in Data Reconciliation

CHAPTER 1

General Introduction

1.6 Quiz

Question 3:

Data reconciliation uses information from

(a) process models.

(b) process measurements.

(c) human common sense.

(d) redundancy in measurements.

Question 4:

If a process variable is measured, then

(a) it is observable.

(b) it is unobservable.

(c) it is maybe redundant.

(d) it is maybe nonredundant.

Basic Concepts in Data Reconciliation

1.7 Suggested Readings

Romagnoli, J.A. and Sanchez, M.C. (2000). "*Data Processing and Reconciliation for Chemical Process Operations*". Academic Press, San Diego.

Narasimhan, S. and Jordache, C. (2000). "*Data Reconciliation & Gross Error Detection, An Intelligent Use of Process Data*". Gulf Publishing, Houston, Texas.

http://www.ris-resolution.com/reconciliation.shtml

http://www.thp.be/reconciliation_nl.html

http://btbjansky.com/prozessdat_e.html

http://www.simsci.com/products/datacon.stm

Basic Concepts in Data Reconciliation

Chapter 2:

Linear Steady-State Data Reconciliation with All Variables Measured

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

The simplest data reconciliation problem occurs in reconciling process flow rates in a plant as illustrated in Figure 1.4. For this example, all the flows are measured in the network. Applying the general data reconciliation algorithm formulated by Equation (1.3), the vector of the raw flow measurements can be written as:
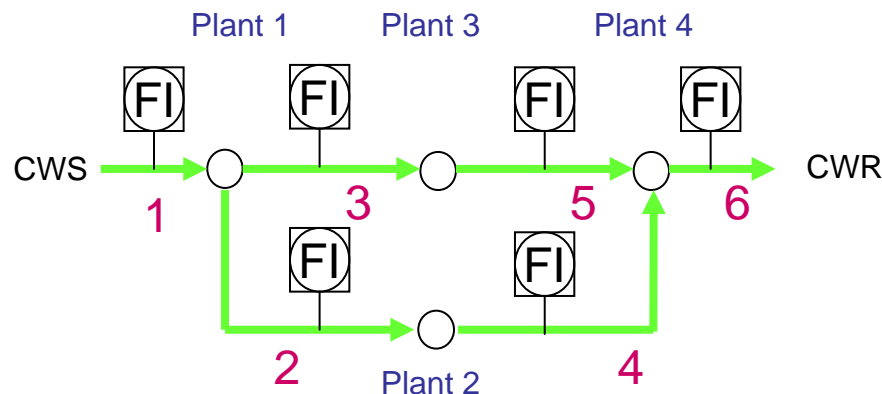
$$\mathbf{y} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \end{bmatrix} = \begin{bmatrix} 110.5 \\ 60.8 \\ 35.0 \\ 68.9 \\ 38.6 \\ 101.4 \end{bmatrix}.$$

Figure 1.4

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

Since we assume that the six measurements are uncorrelated, the variance matrix, **V**, in its diagonal form, can be given as:

$$\mathbf{V} = \begin{bmatrix} 0.6724 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2809 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2116 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5041 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2025 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.44 \end{bmatrix}.$$

(Note: σ^2 is the variance)

Table 1.1

Stream No.	Raw Measurement (kt/h)	Standard Deviation, σ (kt/h)
1	110.5	0.82
2	60.8	0.53
3	35.0	0.46
4	68.9	0.71
5	38.6	0.45
6	101.4	1.20

Basic Concepts in Data Reconciliation

Linear Steady-State Data Reconciliation with All Variables Measured

2.1 Solution to Reconciled Values

The process model constraints in this case are the mass balances around each plant (node) in the network. This is to say that the reconciled values should satisfy the mass balances at each node.

The mass balances around each node can be written as:

Plant 1 $\hat{F}_1 - \hat{F}_2 - \hat{F}_3 = 0$

Plant 2 $\hat{F}_2 - \hat{F}_4 = 0$

Plant 3 $\hat{F}_3 - \hat{F}_5 = 0$

Plant 4 $\hat{F}_4 + \hat{F}_5 - \hat{F}_6 = 0$

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

It is elegant to write the process model constraints (the mass balances) in a compact form, $A\hat{y} = 0$, where

$$A = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}, \quad \hat{y} = \begin{bmatrix} \hat{F}_1 \\ \hat{F}_2 \\ \hat{F}_3 \\ \hat{F}_4 \\ \hat{F}_5 \\ \hat{F}_6 \end{bmatrix},$$

and **0** is a zero-vector.

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

The matrix **A** is called the incidence matrix, where each row represents each node and each column represents each flow stream, respectively. Each element in **A** is either +1, -1 or 0, depending on whether the corresponding flow is an input stream, an output stream, or not associated with this node.

Flow   1   2   3   4   5   6:

Node

$$
\mathbf{A} = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix} \begin{matrix} \leftarrow 1 \\ 2 \\ 3 \\ 4 \end{matrix}
$$

Incidence matrix: **A**

Basic Concepts in Data Reconciliation

CHAPTER 2

Linear Steady-
State Data
Reconciliation
with All
Variables
Measured

2.1 Solution to Reconciled Values

Thus the data reconciliation problem for the cooling water network becomes:

minimizing $\quad J(\hat{\mathbf{y}}) = (\mathbf{y} - \hat{\mathbf{y}})^{\top} \mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$ $\qquad$ (2.1)

subject to $\quad \mathbf{A}\hat{\mathbf{y}} = \mathbf{0}$

The optimization problem of (2.1) can be solved using Lagrange multipliers. The reconciled flow rates are obtained by:

minimizing $\quad J(\hat{\mathbf{y}}) = (\mathbf{y} - \hat{\mathbf{y}})^{\top} \mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}}) - 2\lambda^{\top}\mathbf{A}\hat{\mathbf{y}}$. $\qquad$ (2.2)

where $\quad \boldsymbol{\lambda}^{\top} = [\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4].$

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

The necessary conditions to obtain the minimum of (2.2) are:

$$\frac{\partial J}{\partial \hat{\mathbf{y}}} = -2\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}}) - 2\mathbf{A}^{\top}\lambda = \mathbf{0} \qquad (2.3)$$

$$\frac{\partial J}{\partial \lambda} = \mathbf{A}\hat{\mathbf{y}} = \mathbf{0}$$

Premultiplying each term by the covariance matrix, **V**, in (2.3) yields:

$$\mathbf{y} - \hat{\mathbf{y}} + \mathbf{V}\mathbf{A}^{\top}\lambda = \mathbf{0} \qquad (2.4)$$

Premultiplying each term by the incidence matrix, **A**, in (2.4), and applying $\mathbf{A}\hat{\mathbf{y}} = \mathbf{0}$ yields:

$$\mathbf{A}\mathbf{y} + \mathbf{A}\mathbf{V}\mathbf{A}^{\top}\lambda = \mathbf{0} \qquad (2.5)$$

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

Rearranging equation (2.5) gives:

$$\lambda = -(\mathbf{A}\mathbf{V}^{-1}\mathbf{A}^{\top})^{-1}\mathbf{A}\mathbf{y} \qquad (2.6)$$

Substituting λ in (2.4) and rearranging the equation gives the vector of reconciled data as:

$$\hat{\mathbf{y}} = \mathbf{y} - \mathbf{V}\mathbf{A}^{\top}(\mathbf{A}\mathbf{V}\mathbf{A}^{\top})^{-1}\mathbf{A}\mathbf{y} \qquad (2.7)$$

Equation (2.7) is the basic solution for a linear steady-state data reconciliation problem.

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

It is convenient to use [MATLAB](#) to calculate the reconciled values by Equation (2.7). The solution to the DR problem of the cooling water network is given by the following MATLAB code:

```
*********************************************
y=[110.5;60.8;35.0;68.9;38.6;101.4];
V=[0.6724 0 0 0 0 0;0 0.2809 0 0 0 0;0 0
0.2116 0 0 0;0 0 0 0.5041 0 0;0 0 0 0 0.2025
0;0 0 0 0 0 1.44];
A=[1 -1 -1 0 0 0;0 1 0 -1 0 0;0 0 1 0 -1 0;0 0
0 1 1 -1];
yhat=y-V*A'*inv(A*V*A')*A*y
*********************************************
```

© North Carolina State University, USA   © University of Ottawa, Canada, 2003

Basic Concepts in Data Reconciliation

2.1 Solution to Reconciled Values

The calculation results of the reconciled values for each measurement are listed in Table 2.1. It shows that the reconciled values satisfy mass balances.

Table 2.1: Data reconciliation for a cooling water network

Stream No.	Raw measurement (kt/h)	Standard Deviation, σ (kt/h)	Reconciled Flow (kt/h)	Adjustment (kt/h)
1	110.5	0.82	103.24	-7.26
2	60.8	0.53	65.42	4.62
3	35.0	0.46	37.82	2.82
4	68.9	0.71	65.42	-3.48
5	38.6	0.45	37.82	-0.78
6	101.4	1.20	103.24	1.84

Basic Concepts in Data Reconciliation

CHAPTER 2

Linear Steady-
State Data
Reconciliation
with All
Variables
Measured

2.2 Statistical Properties of Reconciled Values

It is very important that the reconciled values be unbiased. This is to say that the expected values of the reconciled data, $\hat{\mathbf{y}}$, should be equal to the true values of process variables, $\mathbf{x}$.

Recall that the raw measurements can be written as the additive noise model:

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\varepsilon} \tag{2.8}$$

Putting (2.8) into (2.7) gives:

$$\hat{\mathbf{y}} = (\mathbf{x} + \boldsymbol{\varepsilon}) - \mathbf{VA}^\top (\mathbf{AVA}^\top)^{-1} \mathbf{A} (\mathbf{x} + \boldsymbol{\varepsilon}) \tag{2.9}$$

Taking the expected value of (2.9) gives:

$$E(\hat{\mathbf{y}}) = E(\mathbf{x} + \boldsymbol{\varepsilon}) - E[\mathbf{VA}^\top (\mathbf{AVA}^\top)^{-1} \mathbf{A} (\mathbf{x} + \boldsymbol{\varepsilon})] \tag{2.10}$$

Basic Concepts in Data Reconciliation

2.2 Statistical Properties of Reconciled Values

Expanding (2.10) gives:

$$E(\hat{\mathbf{y}}) = E(\mathbf{x}) + E(\boldsymbol{\varepsilon}) - E[\mathbf{VA}^{\top}(\mathbf{AVA}^{\top})^{-1}\mathbf{Ax}] + E(\boldsymbol{\varepsilon})$$

Since $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, and $\mathbf{x}$ is a deterministic variable, $E(\mathbf{x}) = \mathbf{x}$, thus:

$$E(\hat{\mathbf{y}}) = \mathbf{x} - \mathbf{VA}^{\top}(\mathbf{AVA}^{\top})^{-1}\mathbf{Ax}$$

And since $\mathbf{Ax} = \mathbf{0}$ (the true values of the flows satisfy mass balances), $\mathbf{VA}^{\top}(\mathbf{AVA}^{\top})^{-1}\mathbf{Ax} = \mathbf{0}$. Therefore:

$$E(\hat{\mathbf{y}}) = \mathbf{x} \qquad (2.11)$$

Equation (2.11) shows that the reconciled values are unbiased estimates for the linear steady-state reconciliation problem.

Basic Concepts in Data Reconciliation

2.2 Statistical Properties of Reconciled Values

The covariance matrix of the reconciled data can also be obtained. Rewrite Equation (2.7) as:

$$\hat{\mathbf{y}} = [\mathbf{I} - \mathbf{V}\mathbf{A}^{\mathsf{T}}(\mathbf{A}\mathbf{V}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{A}]\mathbf{y} \qquad (2.12)$$

where **I** is the identity matrix. Let $\mathbf{W} = [\mathbf{I} - \mathbf{V}\mathbf{A}^{\mathsf{T}}(\mathbf{A}\mathbf{V}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{A}]$, then (2.12) becomes:

$$\hat{\mathbf{y}} = \mathbf{W}\mathbf{y} \qquad (2.13)$$

From (2.13), the covariance matrix of the reconciled data can be given as:

$$\text{Cov}(\hat{\mathbf{y}}) = \mathbf{W}\text{Cov}(\mathbf{y})\mathbf{W}^{\mathsf{T}} = \mathbf{W}\mathbf{V}\mathbf{W}^{\mathsf{T}} \qquad (2.14)$$

Basic Concepts in Data Reconciliation

2.2 Statistical Properties of Reconciled Values

The covariance matrix calculated by Equation (2.14) for the reconciled flows in the cooling water network is:

$$
\text{Cov}(\hat{y}) =
\begin{bmatrix}
0.1753 & 0.1114 & 0.0639 & 0.1114 & 0.0639 & 0.1753 \\
 & 0.1365 & -0.0251 & 0.1365 & -0.0251 & 0.1114 \\
 & & 0.0890 & -0.0251 & 0.0890 & 0.0693 \\
 & & & 0.1365 & -0.0251 & 0.1114 \\
 & \text{Symm.} & & & 0.0890 & 0.0693 \\
 & & & & & 0.1753
\end{bmatrix}
$$

Note that the covariance matrix of the reconciled flows is symmetric. The diagonal elements are the variances, and the off-diagonal elements are the correlations.

Basic Concepts in Data Reconciliation

2.2 Statistical Properties of Reconciled Values

The standard deviation of the reconciled flows along with the standard deviation of the raw measurements are listed in Table 2.2. It is clear that the reconciled flows have smaller standard deviations and are therefore more precise.

Table 2.2: Variances of reconciled flows

Stream No.	Raw measurement (kt/h)	Standard Deviation, σ (kt/h)	Reconciled Flow (kt/h)	Standard Deviation, σ (kt/h)
1	110.5	0.82	103.24	0.42
2	60.8	0.53	65.42	0.37
3	35.0	0.46	37.82	0.30
4	68.9	0.71	65.42	0.37
5	38.6	0.45	37.82	0.30
6	101.4	1.20	103.34	0.42

Basic Concepts in Data Reconciliation

2.3 Quiz

Question 1:

For an n×n incidence matrix of a flow sheet, rank(A) must be

(a) a full rank of rank n.

(b) less than n.

(c) either rank n, or less than n.

(d) greater than n.

Question 2:

The reconciled data in linear steady-state DR are

(a) more consistent, but less accurate than raw measurements.

(b) more accurate, but less consistent than raw measurements.

(c) more accurate and consistent than raw measurements.

(d) less accurate and consistent than raw measurements.

Basic Concepts in Data Reconciliation

2.3 Quiz

Question 3:

The reconciled data in linear steady-state DR

(a) are biased.

(b) are unbiased.

(c) have a smaller variance than the raw measurements.

(d) have a larger variance than the raw measurements.

Question 4:

The incidence matrix of a flow sheet

(a) is unique.

(b) is not unique.

(c) contains topological information about a flow sheet.

(d) contains the measurement information of a flow sheet.

Basic Concepts in Data Reconciliation

CHAPTER 2

Linear Steady-State Data Reconciliation with All Variables Measured

2.4 Suggested Readings

Romagnoli, J.A. and Sanchez, M.C. (2000). "*Data Processing and Reconciliation for Chemical Process Operations*". Academic Press, San Diego.

Narasimhan, S. and Jordache, C. (2000). "*Data Reconciliation and Gross Error Detection, an Intelligent Use of Process Data*". Gulf Publishing, Houston, Texas.

Mendenhall, W.; Wackerly, D.D. and Scheaffer, R.L. (1990). "*Mathematical statistics with applications*", 4th Ed., PWS, Boston.

Hartfiel, D.J. (2001). "*Matrix theory and applications with MATLAB*". CRC Press, Boca Raton, Fla.

Basic Concepts in Data Reconciliation

Chapter 3:

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

Basic Concepts in Data Reconciliation

CHAPTER 3

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.1 Solutions to Estimates of Measured Variables

In practice, not all flows are measured in a plant due to physical or economical reasons. In this case, we need to develop a DR technique to reconcile the measurements and to estimate unmeasured flow rates as well. The example of the cooling-water network is reused, but with only flows 1, 3 and 5 measured, leaving flows 2, 4 and 6 unmeasured as shown in Figure 3.1.

The problem of data reconciliation with both measured and unmeasured flows can be efficiently solved by the method of Projection Matrix as described in the following slides.

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables
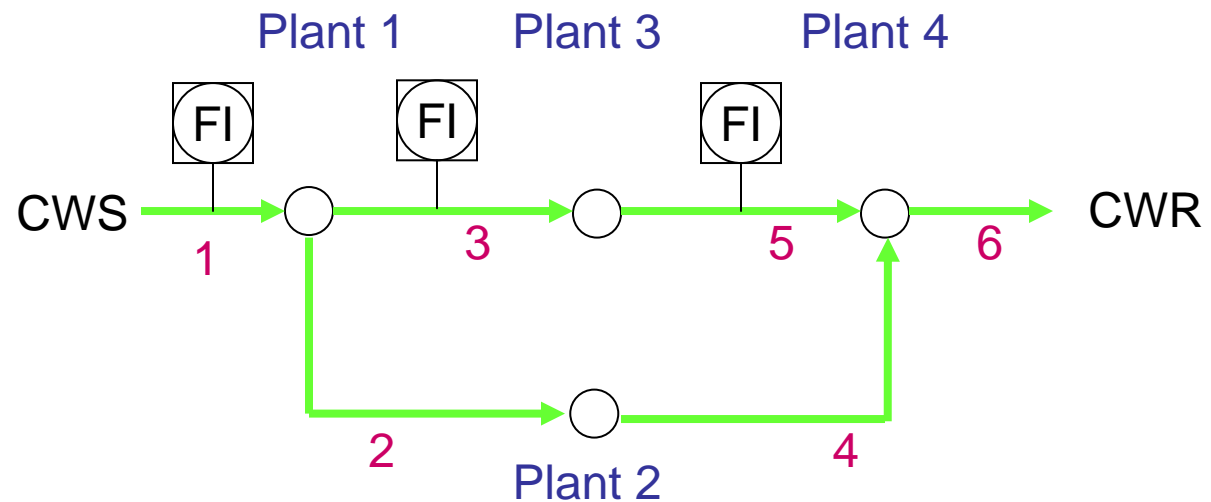
Figure 3.1: Cooling water network with unmeasured flows

First of all, we can partition the incidence matrix of the mass balances in terms of measured and unmeasured flows:

$$\mathbf{A}_y \hat{\mathbf{y}} + \mathbf{A}_z \hat{\mathbf{z}} = \mathbf{0} \qquad (3.1)$$

Basic Concepts in Data Reconciliation

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.1 Solutions to Estimates of Measured Variables

where the columns of $\mathbf{A}_y$ correspond to the measured flows, and those of $\mathbf{A}_z$ correspond to the unmeasured flows. $\hat{\mathbf{y}}$ is the vector of reconciled values for measured flows, and $\hat{\mathbf{z}}$ is the vector of estimates for unmeasured flows. For the example of the cooling water network shown in Figure 3.1, we have:
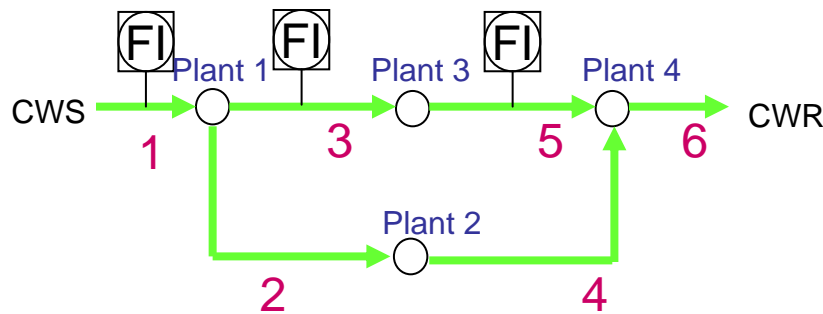
$$\mathbf{A}_y = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \hat{\mathbf{y}} = \begin{bmatrix} \hat{F}_1 \\ \hat{F}_3 \\ \hat{F}_5 \end{bmatrix}$$

$$\mathbf{A}_z = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \qquad \hat{\mathbf{z}} = \begin{bmatrix} \hat{F}_2 \\ \hat{F}_4 \\ \hat{F}_6 \end{bmatrix}$$

Figure 3.1

Basic Concepts in Data Reconciliation

CHAPTER 3

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.1 Solutions to Estimates of Measured Variables

Now we can rewrite the data reconciliation problem as:

minimizing $\quad J(\hat{\mathbf{y}}, \hat{\mathbf{z}}) = (\mathbf{y} - \hat{\mathbf{y}})^{\top}\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$ $\qquad$ (3.2)

subject to $\quad \mathbf{A}_y\hat{\mathbf{y}} + \mathbf{A}_z\hat{\mathbf{z}} = \mathbf{0}$

The solution to the data reconciliation problem (3.2) can be solved by first eliminating the unmeasured flows, $\hat{\mathbf{z}}$, in the constraint equations by pre-multiplying both sides by a projection matrix , $\mathbf{P}$, such that $\mathbf{PA}_z = \mathbf{0}$. Then the data reconciliation problem becomes:

minimizing $\quad J(\hat{\mathbf{y}}) = (\mathbf{y} - \hat{\mathbf{y}})^{\top}\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$ $\qquad$ (3.3)

subject to $\quad \mathbf{PA}_y\hat{\mathbf{y}} = \mathbf{0}$

Basic Concepts in Data Reconciliation

Note: Orthonormal means $Q^TQ = I$

3.1 Solutions to Estimates of Measured Variables

The solution to the optimization problem of (3.3) can be given by Equation (2.7) in which matrix **A** is replaced by matrix $\mathbf{PA_y}$.

$$\hat{\mathbf{y}} = \mathbf{y} - \mathbf{V}(\mathbf{PA}_y)^\top \left[(\mathbf{PA}_y)\mathbf{V}(\mathbf{PA}_y)^\top \right]^{-1} (\mathbf{PA}_y)\mathbf{y} \qquad (3.4)$$

The construction of the projection matrix, **P**, can be performed efficiently using Q-R factorization of matrix $\mathbf{A_z}$.

The statement of the Q-R Theorem (Johnson et al., 1993):

If matrix $\mathbf{A_z}$ (m×n), where m≥n, has columns that are linearly independent (rank($\mathbf{A_z}$) = n), then there is an (m×m) matrix **Q** with orthonormal column vectors such that $\mathbf{A_z} = \mathbf{QR}$, where

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}, \text{ and } \mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

$\mathbf{R}_1$ is an upper triangular and nonsingular matrix with dimension ($n \times n$). $\mathbf{0}$ is a zero matrix with the dimension ($m-n \times n$). $\mathbf{I}$ is an identity matrix.

After the Q-R factorization of matrix $\mathbf{A}_z$, the matrix $\mathbf{Q}$ can be partitioned into two parts as:

$$\mathbf{A}_z = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \tag{3.5}$$

The dimension of $\mathbf{Q}_1$ is ($m \times n$), and $\mathbf{Q}_2$ is ($m \times m-n$).

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables

Premultiplying both sides by Q_2^T in (3.5) yields:

$$Q_2^T A = Q_2^T [Q_1 \quad Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \tag{3.6}$$

Since Q is orthonormal, the matrix Q_2 has the property:

$$Q_2^T [Q_1 \quad Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = [0 \quad I] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = 0$$

Thus $Q_2^T A = 0$.

It is clear that the matrix Q_2^T is the desired projection matrix,

$$P = Q_2^T$$

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables

The Q-R factorization of a matrix A_z can be easily done using a MATLAB command. For the example of the cooling water network, the MATLAB code for Q-R factorization of the matrix A_z is:

```
**********************************
Az=[-1 0 0;1 -1 0;0 0 0;0 1 -1];
[Q,R]=qr(Az)
**********************************
```

The calculation results for the matrices **Q** and **R** from the factorization of A_z by MATLAB are:

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

$$\mathbf{Q} = \begin{bmatrix} -0.707 & -0.408 & 0.577 & 0 \\ 0.707 & -0.408 & 0.577 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0.816 & 0.577 & 0 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1.414 & -0.707 & 0 \\ 0 & 1.225 & -0.816 \\ 0 & 0 & -0.577 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, the matrices **Q** and **R** are decomposed as:

$$\mathbf{Q} = \begin{bmatrix} -0.707 & -0.408 & 0.577 & \vdots & 0 \\ 0.707 & -0.408 & 0.577 & \vdots & 0 \\ 0 & 0 & 0 & \vdots & 1 \\ 0 & 0.816 & 0.577 & \vdots & 0 \end{bmatrix}$$

$$\underbrace{}_{\mathbf{Q}_1} \quad \underbrace{}_{\mathbf{Q}_2}$$

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables

$$\mathbf{R} = \begin{bmatrix} 1.414 & -0.707 & 0 \\ 0 & 1.225 & -0.816 \\ 0 & 0 & -0.577 \\ \cdots & \cdots & \cdots \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} \mathbf{R}_1 \\[1em] \left.\vphantom{1}\right\} \mathbf{0} \end{matrix}$$

Therefore, the projection matrix for this problem is:

$$\mathbf{P} = \mathbf{Q}_2{}^\mathsf{T} = [0 \quad 0 \quad 1 \quad 0]$$

and we have:

$$\mathbf{PA}_y = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}$$

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables

Note that the first element in the matrix $\mathbf{PA}_y$ is zero. This indicates that the measurement F_1 will disappear in the mass balance of the constraint equations in (3.3). The measurement F_1 is nonredundant. This means it can only be evaluated by its measurement.

Now, the data reconciliation becomes to reconcile the two redundant measurements, F_3 and F_5. Rewrite the problem as:

minimizing $\quad J(\hat{\mathbf{y}}) = (\mathbf{y} - \hat{\mathbf{y}})^\top \mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$

subject to $\quad \mathbf{A}\hat{\mathbf{y}} = \mathbf{0}$

where $\mathbf{y} = \begin{bmatrix} F_3 \\ F_5 \end{bmatrix} = \begin{bmatrix} 35.0 \\ 38.6 \end{bmatrix}$, $\mathbf{V} = \begin{bmatrix} 0.2116 & 0 \\ 0 & 0.2025 \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 1 & -1 \end{bmatrix}$.

Basic Concepts in Data Reconciliation

3.1 Solutions to Estimates of Measured Variables

Using Equation (2.7), the reconciled values for F_3 and F_5 are:

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{F}_3 \\ \hat{F}_5 \end{bmatrix} = \begin{bmatrix} 36.84 \\ 36.84 \end{bmatrix}$$

Note that the reconciled values satisfy the mass balance at plant 3. The estimates for the three measured flows by the DR algorithm are:

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{F}_1 \\ \hat{F}_3 \\ \hat{F}_5 \end{bmatrix} = \begin{bmatrix} 110.5 \\ 36.84 \\ 36.84 \end{bmatrix}$$

Basic Concepts in Data Reconciliation

3.2 Solutions to Estimates of Unmeasured Variables

After we obtain the reconciled values (estimates) of the measured flows, $\hat{\mathbf{y}}$, the next step is to estimate the unmeasured flows, $\hat{\mathbf{z}}$, using the information provided by the measured flows and the process models.

From Equation (3.2), the unmeasured flows can be given as:

$$\mathbf{A}_z\hat{\mathbf{z}} = -\mathbf{A}_y\hat{\mathbf{y}} \qquad (3.7)$$

The quantities on the right side of (3.7) are known, so now the problem is to solve the linear equations on the left side. Usually, the number of equations is greater than the number of unmeasured flows. The least-squares technique can then be applied and give the solution of the observable unmeasured flows as:

Basic Concepts in Data Reconciliation

CHAPTER 3

Linear Steady-
State Data
Reconciliation
with Both
Measured and
Unmeasured
Variables

3.2 Solutions to Estimates of Unmeasured Variables

$$\hat{\mathbf{z}} = -(\mathbf{A}_z^T \mathbf{A}_z)^{-1} \mathbf{A}_z^T (\mathbf{A}_y \hat{\mathbf{y}}) \qquad (3.8)$$

For the cooling water network problem, putting the values

$$\mathbf{A}_z = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \text{ and } \mathbf{A}_y \hat{\mathbf{y}} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 110.5 \\ 36.84 \\ 36.84 \end{bmatrix} = \begin{bmatrix} 73.66 \\ 0 \\ 0 \\ 36.84 \end{bmatrix}$$

into (3.8) gives:

$$\hat{\mathbf{z}} = \begin{bmatrix} \hat{F}_2 \\ \hat{F}_4 \\ \hat{F}_6 \end{bmatrix} = \begin{bmatrix} 73.66 \\ 73.66 \\ 110.5 \end{bmatrix}$$

Basic Concepts in Data Reconciliation

3.2 Solutions to Estimates of Unmeasured Variables

The above calculations are carried out here by MATLAB:

```
*************************************************
Az=[-1 0 0;1 -1 0;0 0 0;0 1 -1];
Ay=[1 -1 0;0 0 0;0 1 -1;0 0 1];
yhat=[110.5;36.84;36.84];
zhat=-inv(Az'*Az)*Az'*(Ay*yhat)
*************************************************
```

For convenience, the estimates for the measured and unmeasured flows for the cooling water network are summarized in Table 3.1. Note that the estimates of the flows satisfy mass balances around each plant in the network.

Basic Concepts in Data Reconciliation

3.2 Solutions to Estimates of Unmeasured Variables

Table 3.1: Results of estimation for measured and unmeasured flows in cooling water network

Stream No.	Raw Measurement (kt/h)	Estimated Flow (kt/h)
1	110.5	110.5
2	Unmeasured	73.66
3	35.0	36.84
4	Unmeasured	73.66
5	38.6	36.84
6	Unmeasured	110.5
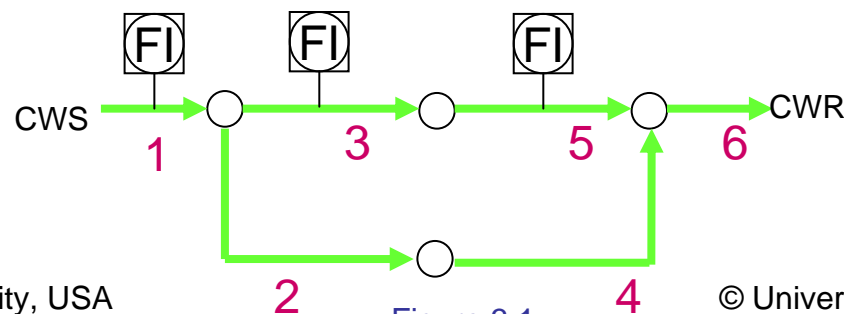
Figure 3.1

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

As stated before, measured variables are either redundant or nonredundant, while unmeasured variables are either observable or nonobservable. The example of the cooling water network shown in Figure 3.1 demonstrated that the measured flows F_3 and F_5 are redundant so that their values can be adjusted. However, the measured flow F_1 is nonredundant, so its value can't be adjusted. On the other hand, all the unmeasured flows are observable since their values can be estimated by the data reconciliation algorithm.

For any process network, the analysis of the observability and redundancy of flow variables can be performed by analyzing the system matrix, **A**, which is the incidence matrix, because the matrix A contains all of the topological information for the network.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

For simplicity, the cooling water network here is cited again. In this example, suppose only flows F_1 and F_6 are measured and the other flows are unmeasured. For this case, the cooling water network is presented in Figure 3.2.
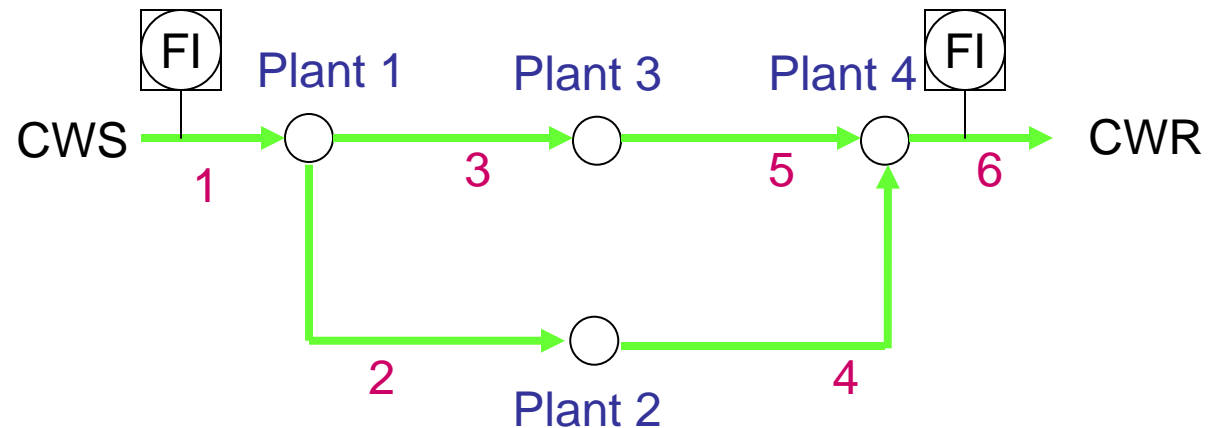
Figure 3.2: Cooling water network with only two flow measurements.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

For convenience, we write the data reconciliation problem of Figure 3.2 as:

minimizing $\quad J(\hat{\mathbf{y}}, \hat{\mathbf{z}}) = (\mathbf{y} - \hat{\mathbf{y}})^\top \mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$

subject to $\quad \mathbf{A}_y \hat{\mathbf{y}} + \mathbf{A}_z \hat{\mathbf{z}} = \mathbf{0}$

where $\quad \mathbf{y} = \begin{bmatrix} F_1 \\ F_6 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} F_2 \\ F_3 \\ F_4 \\ F_5 \end{bmatrix}$. The two partitioned matrices are:

$$\mathbf{A}_y = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & -1 \end{bmatrix} (4 \times 2) , \quad \mathbf{A}_z = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} (4 \times 4) .$$

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

Now, in this case, and in any case where A_z has $n \geqslant m$, it is impossible to split the Q matrix into Q_1 and Q_2 matrices as previously described.

If A_z is 4x4 (mxn), then by the previous rule, Q is 4x4 (mxn), Q_1 is 4x4 (mxm), and Q_2 is 4x0 (mxm-n), which is clearly impossible.

In order to avoid this problem, it must be remembered that $Q_2^T A_z = 0$. The only way of achieving this is to give Q_2 the same number of columns as there are zero rows in the R matrix.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

Back to the cooling water network example, it can be shown that the rank of $\mathbf{A}_z$ is $R(A_z)=3$, but there are 4 unknowns. This means at least one variable in $\hat{\mathbf{z}}$ is undeterminable. In other words, there is at least one flow out of the unmeasured flows that is unobservable.

Performing Q-R factorization of $\mathbf{A}_z$ using MATLAB results in:

$$\mathbf{Q} = \begin{bmatrix} -0.7071 & -0.4082 & -0.2887 & 0.5 \\ 0.7071 & -0.4082 & -0.2887 & 0.5 \\ 0 & 0.8165 & -0.2887 & 0.5 \\ 0 & 0 & 0.8660 & 0.5 \end{bmatrix}$$

Basic Concepts in Data Reconciliation

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.3 Observabilty and Redundancy Analysis

$$R = \begin{bmatrix} 1.4142 & 0.7071 & -0.7071 & 0 \\ 0 & 1.2247 & 0.4082 & -0.8165 \\ 0 & 0 & 1.1547 & 1.1547 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The matrices **Q** and **R**, in this case, are decomposed as:

$$A = QR = \begin{bmatrix} -0.7071 & -0.4082 & -0.2887 & 0.5 \\ 0.7071 & -0.4082 & -0.2887 & 0.5 \\ 0 & 0.8165 & -0.2887 & 0.5 \\ 0 & 0 & 0.8660 & 0.5 \end{bmatrix} \begin{bmatrix} 1.4142 & 0.7071 & -0.7071 & 0 \\ 0 & 1.2247 & 0.4082 & -0.8165 \\ 0 & 0 & 1.1547 & 1.1547 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix}$$

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

where $\mathbf{R}_1$ is the upper triangular matrix having the same rank as matrix $\mathbf{A}_z$. The projection matrix, $\mathbf{P}$, is;

$$\mathbf{P} = \mathbf{Q}_2^{\mathsf{T}} = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}.$$

Then we have:

$$\mathbf{P}\mathbf{A}_y\hat{\mathbf{y}} = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \hat{F}_1 \\ \hat{F}_6 \end{bmatrix} = \hat{F}_1 - \hat{F}_6 = 0$$

The data reconciliation becomes reconciling the flows F_1 and F_6 constrained by a global mass balance around the entire network. The two measurements are redundant.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

Returning to the problem of the observability of unmeasured flows, we know that at least one unmeasured variable is nonobservable in the example of the cooling water network, by analyzing the rank of $\mathbf{A}_z$.

In general, the vector of the unmeasured variables can be partitioned as:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_r \\ \mathbf{z}_{N-r} \end{bmatrix}$$

where r is the rank of $\mathbf{A}_z$ and N is the total number of unmeasured flows. From the rank of $\mathbf{A}_z$, we know that there are at least N-r flows unobservable. The next step is to check the observability of $\mathbf{z}_r$.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

We rewrite the mass balance equations in the form:

$$\begin{bmatrix} \mathbf{A}_y & \mathbf{A}_z \end{bmatrix} \begin{bmatrix} \hat{\mathbf{y}} \\ \hat{\mathbf{z}} \end{bmatrix} = \mathbf{0} \tag{3.9}$$

Premultiplying by matrix $\mathbf{Q}^T$ on both sides of (3.9) gives;

$$\begin{bmatrix} \mathbf{Q}^T\mathbf{A}_y & \mathbf{Q}^T\mathbf{A}_z \end{bmatrix} \begin{bmatrix} \hat{\mathbf{y}} \\ \hat{\mathbf{z}} \end{bmatrix} = \mathbf{0} \tag{3.10}$$

Since $\mathbf{Q}^T = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix}^T = \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix}$, the term $\mathbf{Q}^T\mathbf{A}_y$ in (3.10)

can be written as: $\mathbf{Q}^T\mathbf{A}_y = \begin{bmatrix} \mathbf{Q}_1^T\mathbf{A}_y \\ \mathbf{Q}_2^T\mathbf{A}_y \end{bmatrix}$

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

The term $\mathbf{Q}^T\mathbf{A}_z$ in (3.10) can be written as:

$$\mathbf{Q}^T\mathbf{A}_z = \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1^T\mathbf{Q}_1 & \mathbf{Q}_1^T\mathbf{Q}_2 \\ \mathbf{Q}_2^T\mathbf{Q}_1 & \mathbf{Q}_2^T\mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Because Q is an orthonormal matrix, $\mathbf{Q}_1^T\mathbf{Q}_1 = \mathbf{I}$, $\mathbf{Q}_1^T\mathbf{Q}_2 = \mathbf{0}$, $\mathbf{Q}_2^T\mathbf{Q}_1 = \mathbf{0}$, and $\mathbf{Q}_2^T\mathbf{Q}_2 = \mathbf{I}$. Therefore,

$$\mathbf{Q}^T\mathbf{A}_z = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Now equation (3.10) becomes:

$$\begin{bmatrix} \mathbf{Q}_1^T\mathbf{A}_y & \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{Q}_2^T\mathbf{A}_y & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{y}} \\ \hat{\mathbf{z}}_r \\ \hat{\mathbf{z}}_{N-r} \end{bmatrix} = \mathbf{0} \qquad (3.11)$$

Basic Concepts in Data Reconciliation

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.3 Observabilty and Redundancy Analysis

Equation (3.11) results in two equations:

$$\mathbf{Q}_1^T \mathbf{A}_y \hat{\mathbf{y}} + \mathbf{R}_1 \hat{\mathbf{z}}_r + \mathbf{R}_2 \hat{\mathbf{z}}_{N-r} = \mathbf{0} \tag{3.12}$$

$$\mathbf{Q}_2^T \mathbf{A}_y \hat{\mathbf{y}} = \mathbf{0} \tag{3.13}$$

Note that equation (3.13) is the reduced form of the mass balances by the projection matrix, $\mathbf{Q}_2^T$.

We can rewrite equation (3.12) in terms of $\hat{\mathbf{z}}_r$ as:

$$\hat{\mathbf{z}}_r = -\mathbf{R}_1^{-1} \mathbf{Q}_1^T \mathbf{A}_y \hat{\mathbf{y}} - \mathbf{R}_1^{-1} \mathbf{R}_2 \hat{\mathbf{z}}_{N-r} \tag{3.14}$$

In equation (3.14), the quantities of $\hat{\mathbf{z}}_r$ can be calculated if the rows of the matrix $\mathbf{R}_1^{-1} \mathbf{R}_2$ are zeroes, even though $\hat{\mathbf{z}}_{N-r}$ is unknown (nonobservable).

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

The corollary conclusion from (3.14) can be stated as: the unmeasured variables, z_i, in $\mathbf{z}_r$ are observable if the corresponding elements in the ith row of the matrix $\mathbf{R}_1^{-1}\mathbf{R}_2$ are all zeroes.

For the example of the cooling water network shown in Figure 3.2, the vector of the unmeasured flows is decomposed as:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_r \\ \mathbf{z}_{N-r} \end{bmatrix} = \begin{bmatrix} F_2 \\ F_3 \\ F_4 \\ F_5 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}F_2\\F_3\\F_4\end{matrix}}\right\} \mathbf{z}_r \\ \left.\vphantom{F_5}\right\} \mathbf{z}_{N-r} \end{matrix}$$

since F_5 is nonobservable. $\mathbf{R}_1^{-1}\mathbf{R}_2$ is calculated as:

Basic Concepts in Data Reconciliation

CHAPTER 3
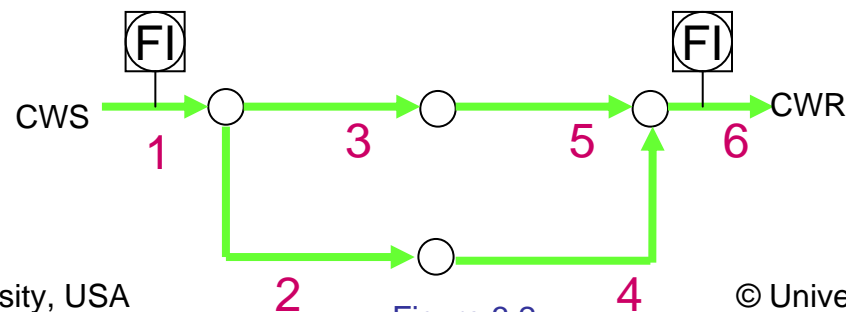
Linear Steady-
State Data
Reconciliation
with Both
Measured and
Unmeasured
Variables

3.3 Observabilty and Redundancy Analysis

$$\mathbf{R_1^{-1}R_2} = \begin{bmatrix} 1.4142 & 0.7071 & -0.7071 \\ 0 & 1.2247 & 0.4082 \\ 0 & 0 & 1.1547 \end{bmatrix}^{T} \begin{bmatrix} 0 \\ -0.8165 \\ 1.1547 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

This shows that there is no zero-row in $\mathbf{R_1^{-1}R_2}$, so that all three unmeasured flows, F_2, F_3, and F_4 are also nonobservable.

Actually, from Figure 3.2 , it is clear that all the unmeasured flows can't be evaluated. The above analysis seems unnecessary. However, for complex process networks, the advantages of the above analysis will be obvious.

Figure 3.2

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

The Q-R factorization method introduced is also valid when A_z is of the dimension (m×n), where m<n. Sometimes, the calculated matrix **R** from the factorization of A_z has zero-rows that are located above non-zero rows. For example:

$$A_z = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \qquad (3.15)$$

The Q-R factorization of A_z results in:

$$Q = \begin{bmatrix} -0.7071 & -0.7071 & 0 \\ 0.7071 & -0.7071 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad R = \begin{bmatrix} -1.4142 & 1.4142 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

The elements in the second row of **R** are all zeroes. In this case, the matrix $\mathbf{A}_z$ needs a column permutation such that:

$$\mathbf{A}_z \Pi = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where Π is a permutation matrix having the property $\Pi^T = \Pi = \Pi^{-1}$. For this example, the permutation is:

$$\mathbf{A}_z \Pi = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

MATLAB has a very useful command that will calculate both the permutation matrix and the permuted Q and R matrices. For this example:

```
*********************************************
Az=[1 -1 0 0;-1 1 0 0;0 0 1 -1];
[Q,R,E]=qr(Az)
*********************************************
```

Multiplying the A_z matrix by the permutation matrix is unnecessary here, as it is automatically done by the [Q,R,E] command to produce the altered Q and R matrices.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

In any case, the Q-R factorization of $\mathbf{A}_z\Pi$ results in:

$$\mathbf{A}_z\Pi = \begin{bmatrix} -0.7071 & 0 & \vdots & -0.7071 \\ 0.7071 & 0 & \vdots & -0.7071 \\ 0 & -1 & \vdots & 0 \end{bmatrix} \begin{bmatrix} -1.4142 & 0 & \vdots & 1.4142 & 0 \\ 0 & -1 & \vdots & 0 & 1 \\ 0 & 0 & \vdots & 0 & 0 \end{bmatrix}$$

In general, the mass balances in the data reconciliation problem of (3.2) can be written in the form:

$$\mathbf{A}_y\hat{\mathbf{y}} + \mathbf{A}_z\Pi\Pi^\mathsf{T}\hat{\mathbf{z}} = \mathbf{0} \qquad \text{(Note: } \Pi\Pi^\mathsf{T}=\mathbf{I}\text{)}$$

$$\mathbf{A}_y\hat{\mathbf{y}} + (\mathbf{A}_z\Pi)(\Pi^\mathsf{T}\hat{\mathbf{z}}) = \mathbf{0}$$

where $\mathbf{A}_z\Pi$ is used to find the projection matrix, $\mathbf{P}$, and Π^T permutes the unmeasured variables in the $\hat{\mathbf{z}}$ vector.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

The permutation matrix, Π, enables an easy classification of the unmeasured variables, as:

$$\Pi^\top \hat{\mathbf{z}} = \begin{bmatrix} \hat{\mathbf{z}}_r \\ \hat{\mathbf{z}}_{N-r} \end{bmatrix}$$

where the variables in the subset $\hat{\mathbf{z}}_{N-r}$ are the minimum number of unmeasured variables that need to be measured for the network to satisfy the observability condition.

Now our study returns to the redundancy analysis of the measured variables. For this problem, the matrix $\mathbf{Q}_2^\top \mathbf{A}_y$ in Equation (3.13) contains the information of the redundancy of the measured process variables.

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

The zero-columns in $\mathbf{Q}_2^T\mathbf{A}_y$ correspond to variables in $\hat{\mathbf{y}}$ that will not participate in the data reconciliation, so they are nonredundant. The remaining non-zero columns in $\mathbf{Q}_2^T\mathbf{A}_y$ correspond to redundant measurements in $\hat{\mathbf{y}}$.

Actually, the above statements have been justified by the cooling water example shown in Figure 3.1.

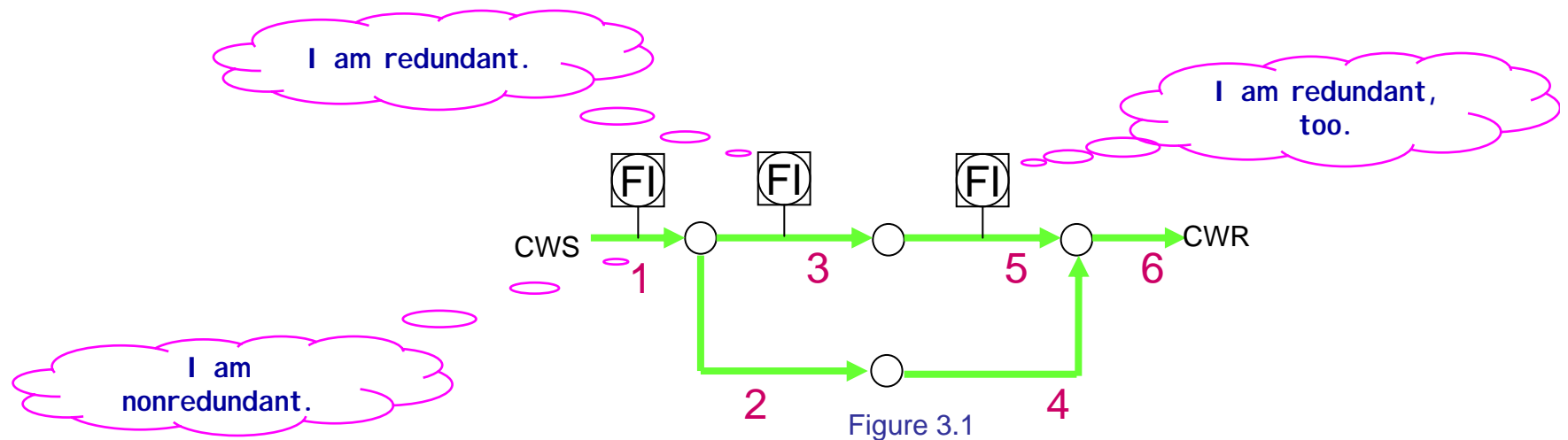Figure 3.1

Basic Concepts in Data Reconciliation

3.3 Observabilty and Redundancy Analysis

The problem of linear steady-state data reconciliation with both measured and unmeasured variables can be efficiently solved using the projection matrix method. This technique is summarized in the following steps.

Step 1: Decompose the system matrix, $\mathbf{A}$, in terms of $\mathbf{A}_y$ and $\mathbf{A}_z$, which correspond to measured and unmeasured variables.

Step 2: Check the rank of $\mathbf{A}_z$.

Step 3: If $R(\mathbf{A}_z) \geq N$, where N is the number of unmeasured variables, then all unmeasured variables are observable. Conduct the data reconciliation formulated by Equation (3.3), and estimate the unmeasured variables using Equation (3.8). Otherwise go to Step 4.

Basic Concepts in Data Reconciliation

CHAPTER 3

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.3 Observabilty and Redundancy Analysis

Step 4: If R($\mathbf{A}_z$) = r < N, then at least (N-r) variables can't be estimated from the available information. Find the permutation matrix Π, such that $\mathbf{A}_z\Pi$ is factorized as:

$$\mathbf{A}_z\Pi = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Step 5: Get the projection matrix, $\mathbf{P} = \mathbf{Q}_2^T$. Proceed with the data reconciliation using Equation (3.3). Only redundant measured variables participate in the data reconciliation. The nonredundant measurements are identified by the matrix $\mathbf{Q}_2^T\mathbf{A}_y$. Obtain the estimates for the measured variables.

Step 6: Calculate the unmeasured variables using Equation (3.14). Only the unmeasured variables in $\hat{\mathbf{z}}_r$ corresponding to zero-rows in the matrix $\mathbf{R}_1^{-1}\mathbf{R}_2$ can be calculated.

Basic Concepts in Data Reconciliation

CHAPTER 3

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.4 Quiz

Question 1:

If $\mathbf{A}_z$ has a full rank, then the unmeasured variables

(a) are all observable.

(b) are all nonobservable.

(c) have some that are observable and some that are not.

(d) and measured variables are redundant.

Question 2:

If the rank of $\mathbf{A}_z$ is $r < N$, where N is the total number of unmeasured variables, then

(a) at least (N-r) variables are nonobservable.

(b) there are exactly (N-r) variables nonobservable.

(c) there are r variables observable.

(d) there are r variables nonobservable.

Basic Concepts in Data Reconciliation

3.4 Quiz

Question 3:

Comparing the two Equations (3.8) and (3.14) ,

(a) they are equivalent.

(b) equation (3.8) is used only when all variables are observable.

(c) equation (3.14) can always be applied whether all variables are observable or not.

(d) equation (3.14) can never be applied whether all variables are observable or not.

Question 4:

For the case where the unmeasured variables can be calculated by Equation (3.8), show that the expected values of the estimates are

$$E(\hat{\mathbf{z}}) = -(\mathbf{A}_z^T \mathbf{A}_z)^{-1} \mathbf{A}_z^T \mathbf{A}_y \mathbf{x}_y$$

where $\mathbf{x}_y$ is the true values of measured variables.

Basic Concepts in Data Reconciliation

CHAPTER 3

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.5 Suggested Readings

Crowe, C.M.; Campos, G. and Hrymak, A. (1983). Reconciliation of process flow rates by matrix projection, Part I: Linear case. *AIChE J*. 29, 881-888.

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

Johnson, L.W.; Riess, R.D. and Arnold, J.T. (1993). "*Introduction to Linear Algebra*". 3th Ed., Addison-Wesley, Reading, Massachusetts.

Romagnoli, J.A. and Sanchez, M.C. (2000). "*Data Processing and Reconciliation for Chemical Process Operations*". Academic Press, San Diego.

Basic Concepts in Data Reconciliation

CHAPTER 3

Linear Steady-State Data Reconciliation with Both Measured and Unmeasured Variables

3.5 Suggested Readings

Sanchez, M. and Romagnoli, J. (1996). Use of orthogonal transformations in data classification – reconciliation. *Computers Chem. Engng.*, 20, 483-493.

Sen, A. and Srivastava, M. (1990). "*Regression analysis : theory, methods and applications*". Springer-Verlag, New York.