XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6−11, 2009, Lisbon, Portugal

DATA RECONCILIATION AND THE SINGULAR VALUE DECOMPOSITION

Christos L. Mitsas

Hellenic Institute of Metrology (EIM), Sindos 57022, Thessaloniki, Greece, chris.mitsas@eim.gr

Abstract − The steady state data reconciliation problem is approached via a geometrical picture of its model and measurement abstract spaces. By completely utilizing the structure of the problem constraint matrix, via its singular value decomposition (SVD), data adjustment is accomplished and redundancy and observability conditions are formulated. As an example, the method is applied to a small network of liquid flowmeters in order to ascertain the reliability of the measurement results.

Keywords (up to three): data reconciliation, SVD

1. INTRODUCTION

Data reconciliation has been used extensively during the past 30 years as a tool for control of industrial processes utilizing the measurements which are collected during process monitoring [1, 2]. By their nature, measurements contain inaccurate information which manifests itself as measurement error and which can be due to non-ideal sensor behavior or to the process itself. Maximizing the efficient use of the information which is available through the monitoring of the process is dependent to a large extent on the effective cancellation of these errors through data conditioning. Given the fact that the reliability of data collected during process monitoring is of the utmost importance in decision making and performance optimization, the use of a statistical technique such as data reconciliation is very appealing. Data reconciliation relies on the existence of redundancy to perform data adjustment based on a least squares criterion, whilst satisfying the model constraints representing physical laws underpinning the relationships between the measured variables. Furthermore, in situations where it is not feasible to measure all process variables it can furnish unmeasured variable estimates through the model constraints, provided they are observable [3].

In the present work, the data reconciliation problem is analyzed by employing the "fundamental theorem of algebra" [4]. This approach could present certain advantages over other matrix decomposition or graph theoretic methods [3] by providing a clear geometrical picture and making more tractable complex problems by reducing their dimensionality as well as by easily identifying and providing estimates of unmeasured but observable variables. In order to apply the above, the singular value decomposition (SVD) [5] of the constraint matrix is used which decomposes the problem into vector subspaces which have the convenient property of being orthogonal complements of each other. As an example, the method is applied to adjust process data from a small liquid flowmeter network under different measurement conditions.

2. THEORY

Given a process network with X_i, i = 1…n variables, the model describing their measured estimates y_i is

$$\mathbf{y} = \mathbf{X} + \boldsymbol{\varepsilon} \qquad (1)$$

where **X**, **y** and **ε** are nx1 matrices representing the variable, measurement estimate and random error vectors with $E(\boldsymbol{\varepsilon}) = 0$, $E(\boldsymbol{\varepsilon\varepsilon}^T) = \boldsymbol{\Sigma}$ a nxn covariance matrix. If the process model is that of generalized mass balance equations at the m nodes of the network, assuming the absence of sources or sinks, it is represented by

$$\mathbf{AX = 0} \qquad (2)$$

where **A** is a mxn constraint matrix. The classical data reconciliation problem reduces to the estimation of the variables **X**, which minimize the RSS error by employing the criterion of weighted least squares while simultaneously satisfying the constraints, i.e.,

$$\min_{X} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \min_{X} [(\mathbf{y - X})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y - X}) + 2\boldsymbol{\lambda}^T \mathbf{AX}] \qquad (3)$$

where $\boldsymbol{\lambda}^T$ is a 1xm vector of Lagrange multipliers [6].

Alternatively, the problem can be formulated without resorting to Lagrange multipliers, by considering that (2) only allows estimates which belong to the null space *N(A)*, of **A**, which is spanned by the solutions of **AX** = **0** and is of dimension n-r, where r is the column (row) rank of **A** [7]. Consequently, (3) can be written as

$$\min_{\beta} (\mathbf{y - W\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y - W\beta}) \qquad (4)$$

where **W** is the nx(n-r) matrix whose columns are the basis vectors of *N(A)* and the vector **β** (n-r)x1 includes the estimates of the model parameters. The most reliable method of analyzing the structure of **A** is through its SVD, e.g., the mxn matrix **A** is decomposed as $\mathbf{A} = \mathbf{UDV}^T$ where the orthogonal matrices **U** and **V** are mxm and nxn and are composed of the eigenvectors of $\mathbf{AA}^T$ και $\mathbf{A}^T\mathbf{A}$ respectively. The mxn matrix **D** has its non-negative elements (the singular values which are equal to the square roots of the non-zero eigenvalues of $\mathbf{A}^T\mathbf{A}$) in descending order which occupy the first r positions of the diagonal. Essentially, the SVD constructs orthonormal bases for the fundamental subspaces of $\mathbf{R}^m$ and $\mathbf{R}^n$ [4], i.e.,

$$\mathbf{A} = (\underbrace{\mathbf{U}_1^{mxr}}_{basisR(A)}, \underbrace{\mathbf{U}_2^{mx(m-r)}}_{basisN(A^T)}) \begin{vmatrix} \mathbf{D}^{rxr} & \mathbf{0}^{rx(n-r)} \\ \mathbf{0}^{(m-r)xr} & \mathbf{0}^{(m-r)x(n-r)} \end{vmatrix} \mathrm{x} \; \mathrm{x}(\underbrace{\mathbf{V}_1^{nxr}}_{basisR(A^T)}, \underbrace{\mathbf{V}_2^{nx(n-r)}}_{basisN(A)})^T \tag{5}$$

From (5) it becomes evident that a feasible solution of problem (4) only exists if $N(A) \neq \{0\}$, i.e., if the mxn matrix **A** is either rank deficient or full row rank (when m < n). In the former case the solution is not complete since it is also true that $N(A^T) \neq \{0\}$, i.e., the mapping $\mathbf{R}^m \rightarrow \mathbf{R}^n$ is singular.

In situations of partially measured systems, i.e., when all the variables X_i are not measured, the structure of the constraint matrix can be exploited so that all or some of the unmeasured variables can be estimated and as such be observable [3]. Since some of the constraints will be used for this purpose, this implies that the full constraint set might not be available for adjustment of all measured variables, in other words, some of them will be non-redundant. Cases of such partially measured systems can be handled efficiently by separating the constraint set of (2) into two parts [8], as

$$\begin{vmatrix} \mathbf{A_X} & \mathbf{A}_\xi \end{vmatrix} \cdot \begin{vmatrix} \mathbf{X'} \\ \xi \end{vmatrix} = \mathbf{0} \tag{6}$$

X' and ξ being the vectors of the measured and unmeasured variables respectively, $\mathbf{A_X}$ and $\mathbf{A}_\xi$ being the corresponding mx(n-s) and mxs constraint matrices and s the number of unmeasured variables.

Since $\mathbf{A}_\xi \in R(A_\xi)$, an mxm matrix **P** which projects onto its left null space, $N(A_\xi^T)$, applied to (6) will yield the relation

$$\mathbf{P\,A_X X' = A_r X' = 0} \tag{7}$$

eliminating the constraints involving unmeasured variables. $\mathbf{A_r}$ is a reduced constraint matrix linking only measured variables. Furthermore, the projection matrix **P**, in terms of the SVD of $\mathbf{A}_\xi$, is written as

$$\mathbf{P = I_{mxm} - U_{\xi 1} U_{\xi 1}^T} \tag{8}$$

where $\mathbf{I_{mxm}}$ is the unit matrix and $\mathbf{U}_{\xi 1}$ is the orthogonal matrix consisting of the basis vectors of $R(A_\xi)$. The expression above is non-zero only when $N(A_u^T) \neq \{0\}$, i.e., when matrix $\mathbf{A}_\xi$ is either rank deficient or full column rank (when m > s). With this projection matrix, it follows from (7) that the reduced constraint matrix consists of columns of $\mathbf{A_X}$ after subtracting their components on $R(A_\xi)$. This in turn implies that if a column of $\mathbf{A_X}$ is entirely in the $R(A_\xi)$ then the corresponding variable cannot be reconciled, i.e., it is non-redundant. This is a consequence of the lack of available information since it is only linked to unmeasured variables via the constraints. The estimation of the redundant variables can proceed via (4) where the matrix **W** consists of the basis vectors of $N(A_r)$.

Conditions of observability on unmeasured variables can now be established through (6), which, after adjustment of redundant variables has been accomplished, can be rewritten as

$$-\left(\mathbf{A}_\xi^T \mathbf{A}_\xi\right)^{-1} \mathbf{A}_\xi^T \mathbf{Z} = \hat{\xi} \tag{9}$$

where **Z** is an mx1 known vector and $\hat{\xi}$ is the sx1 vector of estimates of the unmeasured variables. From the above it follows that all estimates can be obtained if $N(A_\xi) = \{0\}$, i.e., $\mathbf{A}_\xi$ is full column rank (when m > n), so that the positive definite matrix $\mathbf{A}_\xi^T\mathbf{A}_\xi$ is invertible. When $\mathbf{A}_\xi$ is either rank deficient or full row rank (when m < n) (9) cannot formally be solved. Fortunately, the use of a generalized inverse [9] via its SVD will yield a minimum norm solution $\hat{\xi}^*$ as

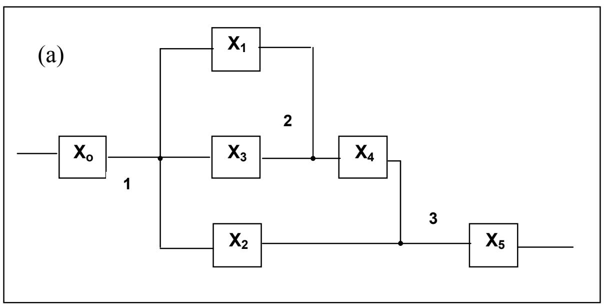
$$-\mathbf{V}_\xi \mathbf{D}^+ \mathbf{U}_\xi^T \mathbf{Z} = \hat{\xi}^* \tag{10}$$

by eliminating contributions which originate from $N(A_\xi)$. The components of this solution vector which correspond to the columns of $\mathbf{A}_\xi$ for which $\mathbf{V}_{\xi 2}^T\mathbf{A}_\xi = \mathbf{0}$ are estimates of observable variables.

3. RESULTS

3.1. Reconciliation with all variables measured

The methodology is applied to a simple flowmeter network, with all variables measured, depicted in figure 1a, with figure 1b showing the corresponding constraint matrix resulting from mass balance at the 3 nodes of the network.



(b)

variables	0	1	2	3	4	5
nodes						
1	1	-1	-1	-1	0	0
2	0	1	0	1	-1	0
3	0	0	1	0	1	-1

Fig. 1. (a) The simple flowmeter network considered and (b) the constraint matrix **A** which is of full row rank (rank (**A**) = 3).

Furthermore, the first two columns of table 2 show the actual flowmeter indications and the corresponding expanded uncertainties (k=2) of the measurements.

From the SVD of the constraint matrix, as implemented on MATLAB,

$$\mathbf{A} = \mathbf{U}_1^{3x3}\left|\mathbf{D}^{3x3} \quad \mathbf{0}^{3x3}\right|(\mathbf{V}_1^{6x3}, \mathbf{V}_2^{6x3})^{\mathrm{T}} \qquad (10)$$

it can be concluded that:

1. there exist 3 non-zero singular values, i.e., the matrix is of rank = 3
2. the left nullspace is empty, $N(A^T) = 0$, since the matrix $\mathbf{U_2}$ does not exist. Consequently the solution which will result from weighted least squares fitting will be complete.
3. the nullspace $N(\boldsymbol{A})$ consists of 3 basis vectors (matrix $\mathbf{V_2}^{6x3}$) which can be used for the solution by performing: $\min_{\beta} (\mathbf{y} - \mathbf{V}_2^{6x3}\boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{V}_2^{6x3}\boldsymbol{\beta})$

The data reconciliation results are shown in table 1 along with the associated expanded uncertainties (95% confidence level). The uncertainties were estimated from the covariance matrix, cov($\boldsymbol{\beta}$), of the parameters, $\boldsymbol{\beta}$, and subsequent propagation of uncertainty [6], via the relation

$$\mathrm{cov}(\hat{\mathbf{y}}) = \mathbf{V}_2(\mathbf{V}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{V}_2)^{-1}\mathbf{V}_2^{\mathrm{T}} \qquad (11)$$

where the dimensions of $\mathbf{V_2}$ have been left out for clarity.

Table 1. Data reconciliation results on the network of figure 1a.

y (L)	U(y) (k=2)	$\hat{y}$ (L)	u($\hat{y}$) (k=1)	$((y-\hat{y})/u)^2$	$\hat{y}$ -y
20,45	0,82	20,85	0,23	0,951	0,40
5,31	0,31	5,30	0,13	0,006	-0,01
9,74	0,49	9,54	0,21	0,635	-0,20
6,02	0,32	6,01	0,14	0,006	-0,01
11,47	0,49	11,30	0,15	0,454	-0,17
20,39	1,45	20,85	0,23	0,402	0,46

The "goodness of fit" is estimated by the index χ^2. The statistical significance test to be conducted is Prob($\chi^2(\nu) \geq \chi^2_{obs}$) < 5%, where ν the degrees of freedom, which in the particular example are ν=3 and χ^2_{obs} the squared sum of the weighted residuals. From the calculation we obtain χ^2 (5%, 3)=2,60 > 0,82 = χ^2_{obs} hence the null hypothesis of no systematic errors is accepted at the 5% level of significance.

3.2. Reconciliation including non-measured variables

Let us re-examine the flowmeter network depicted in figure 1a but now because of metering cost reduction the variables X_1 and X_3 are not measured even though an estimate of their values is still required. Following the treatment of section 2, the constraint matrix of fig. 1b is partitioned into two parts, the one corresponding to the unmeasured variables being

$$A_\xi = \begin{vmatrix} -1 & -1 \\ 1 & 1 \\ 0 & 0 \end{vmatrix}$$

which is obviously rank deficient, implying in turn that $N(A_\xi) \neq \{0\}$, hence the unmeasured variables are not observable. The SVD of this matrix is $\mathbf{A}_\xi = \mathbf{U}_{\xi1}^{3x1}\mathbf{U}_{\xi2}^{3x2}\mathbf{D}_\xi^{2x2}(\mathbf{V}_{\xi1}^{2x1}, \mathbf{V}_{\xi2}^{2x1})^{\mathrm{T}}$ where the smallest singular value of $\mathbf{D}_\xi^{2x2}$ is zero. This is in agreement with the non-observability rule determined by graph theoretic methods concerning loops consisting solely of unmeasured flow streams [2, 3].

Additionally, from the SVD it is seen also that $N(A_\xi^T) \neq \{0\}$, signifying that at least some redundancy exists among the measured variables. Application of the projection $\mathbf{P}$ on the matrix of measured variables results in the reduced constraint matrix

$$A_r = \begin{vmatrix} 0,5 & -0,5 & -0,5 & 0 \\ 0 & 1 & 1 & -1 \end{vmatrix}$$

which is of full row rank, its SVD being $\mathbf{A}_r = \mathbf{U}_{r1}^{2x2}\mathbf{D}_r^{2x2}(\mathbf{V}_{r1}^{4x2}, \mathbf{V}_{r2}^{4x2})^{\mathrm{T}}$. The results of the adjustment by employing its null space basis vectors are shown in table 2.

Table 2. Data reconciliation results when variables X_1 and X_3 are unmeasured.

X_i	y (L)	U(y) (k=2)	$\hat{y}$ (L)	u($\hat{y}$) (k=1)	$\hat{y}$ -y
0	20,45	0,82	20,49	0,41	0,04
2	9,74	0,49	9,64	0,23	-0,10
4	11,47	0,49	11,37	0,23	-0,10
5	20,39	1,45	21,11	0,36	0,72

In a seemingly similar situation, assume that the variables X_2 and X_4 are not measured. Note that now the variables are part of a loop consisting of both measured and unmeasured flow streams. The constraint matrix corresponding to the unmeasured variables now is

$$A_\xi = \begin{vmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{vmatrix}$$

with an SVD $\mathbf{A}_{\xi u} = \mathbf{U}_{\xi1}^{3x2}\mathbf{U}_{\xi2}^{3x1}\mathbf{D}_\xi^{2x2}(\mathbf{V}_{\xi1}^{2x2})^{\mathrm{T}}$ and the smallest singular value of $\mathbf{D}_\xi^{2x2}$ non-zero. Since $N(A_\xi) = \{0\}$ then both the unmeasured variables are observable and can be estimated through (9). In addition $N(A_\xi^T) \neq \{0\}$ suggesting that some degree of redundancy exists for the measured variables which can be used to produce adjusted estimates through the reduced constraint matrix

$A_r = |\ 0{,}33 \quad 0 \quad 0 \quad -0{,}33\ |$

It is evident from the zero's in the second and third columns that the corresponding variables are non-redundant because they belong to *R(A*$_\xi$*)*. The SVD of this matrix is $\mathbf{A_r = U_{r1}^{1x1} | D_r^{1x1} \quad 0^{1x3} | (V_{r1}^{4x1}, V_{r2}^{4x3})^T}$, yielding adjusted variable values by employing its null space basis vectors as shown in table 3. In the same table are included the estimates of the observable variables also. It is interesting to notice that the values for variables X_1 and X_3 have not been reconciled in the process since they do not appear in the reduced constraint matrix. The uncertainties of the estimates of the observable variables were determined via the relation

$$\mathbf{cov(\hat{\xi}) = V_\xi D_\xi^{+} (U_\xi^T A_X) cov(\hat{y}) (U_\xi^T A_X)^T (V_\xi D_\xi^{+})^T} \quad (12)$$

where $\mathbf{cov(\hat{y})}$ is the covariance matrix of the estimates of the measured variables.

Table 3. Data reconciliation results when variables X_2 and X_4 are unmeasured.

X_i	y (L)	U(y) (k=2)	$\hat{y}$ (L)	u($\hat{y}$) (k=1)	$\hat{y}$ -y
0	20,45	0,82	20,44	0,36	-0,01
1	5,31	0,31	5,31	0,16	--
2	--	--	9,11	0,42	--
3	6,02	0,32	6,02	0,16	--
4	--	--	11,33	0,22	--
5	20,39	1,45	20,44	0,36	0,05

4. DISCUSSION

The justification of the use of the SVD in data reconciliation problems warrants some comments, due to its computational intensity and given the fact that other algorithms such as the QR decomposition are effectively implemented in existing software packages. In fact the computational cost of the SVD is, at least, to a degree balanced by the reduced dimensionality of the transformed estimation problem. Thus instead of having to estimate n parameters according to the measurement model (1), under the constraints of (2), only n-r parameters are determined by using the null space basis vectors as determined through the SVD of the constraint matrix. Furthermore, the complete decomposition by the SVD of model and measurement spaces in the case of partially measured systems affords at little extra cost a classification of variables as observable and redundant. In fact, this is the great advantage of the method which through a clear picture of the geometry of the data adjustment can diagnose potential ill-posedness problems and at least point to the direction of how to overcome them.

Finally, it should be mentioned that the uncertainty estimates of the reconciled data, as shown in tables 1 to 3, are quoted to a confidence level well below that of 95%. The reason for this is the limited number of degrees of freedom of the estimates in the example situations presented which produce unrealistically large coverage factors for the above confidence level.

5. CONCLUSIONS

Based on a study of the null spaces of all constraint matrices which can appear in steady state data reconciliation problems, algebraic observability and redundancy conditions have been formulated. The singular value decomposition of the model constraint matrix coupled with the notion of the four fundamental subspaces of a linear transformation provide a clear geometric picture of its structure enabling null space methods to be readily applied.

REFERENCES

[1] **J. Romagnoli, G. Stephanopoulos**, *Chem. Eng. Science* **35**, pp.**1067-1081, 1980.**

[2] **R.S.H. Mah, G.M. Stanley and D.W. Downing**, *Ind &Eng. Chem. Proc. Des. Dev.* **15**, pp.**175-183, 1976**.

[3] **S. Narasimhan, C. Jordache**, *"Data Reconciliation and Gross Error Detection"*, **Gulf Pub. Co. 2000.**

[4] **G. Strang**, *"Linear Algebra and its Applications"*, **3rd Ed., Harcourt Brace Jovanovich, San Diego 1988.**

[5] **G. Golub, C. Van Loan,** *"Matrix Computations"*, **3rd ed. Johns Hopkins University Press 1996.**

[6] **I. Lira,** *"Evaluating the Measurement Uncertainty"*, **IOP Publishing Ltd, Bristol 2002.**

[7] **C. Lawson, R. Hanson,** *"Solving Least Squares Problems"*, **Classics in Applied Mathematics 15, SIAM 1995**.

[8] **C.M. Crowe,** *Chem. Eng. Sci.* **44**, **pp.2909-2917, 1989**.

[9] **R. Penrose**, *Proc. Camb. Phil. Soc.* **51**, pp.**406-413, 1955.**