Test of Maximum Power for Detection of Gross Errors in Process Constraints

Cameron M. Crowe
Department of Chemical Engineering
McMaster University
Hamilton, Ontario
Canada L8S 4L7

Introduction

Since the measurements of concentrations and flow rates in a process at steady state are subject to random error, they do not, in general, obey the laws of conservation and other appropriate constraints. The reconciliation of these measurements with the set of constraints is an important step, both in monitoring process performance, and in modeling.

Unfortunately, there may be errors present that are not random. These gross errors must be identified and either deleted or corrected before carrying out the reconciliation, since otherwise, they will corrupt the results. In order to detect any gross errors, several statistical tests have been proposed, based on the assumption that the measurements are a random sample of the true values at steady state.

The global chi-square test (Reilly and Carpani, 1963) compares the value of the optimal objective function to an appropriate tabulated chi-square value. If the former exceeds the latter, the test is violated, indicating the presence of gross errors.

For the linear case, where the measurements consist solely of species flow rates, a test statistic for the individual constraints was defined by Reilly and Carpani (1963). Mah and Tamhane (1982), and Crowe et al. (1983) proposed an analogous test statistic for the individual measurements of species flow. In each case, the statistic is normally distributed with mean zero and unit variance (unit normal variate), and its value is compared with the threshold value from tables. Similar test statistics for the bilinear case, where concentration and total flow measurements are arbitrarily located, were defined by Crowe (1986).

Tamhane (1982) derived a measurement test of maximum power (*MP*) in the sense that the test has a greater probability of correctly finding a single gross error in a particular measurement than would any other test based on a linear combination of the measurements. It is the aim of this paper to extend the concept of maximum power to the constraint test.

Maximum Power (*MP*) Test on Constraints

The *MP* test on constraints can be developed for any m-dimensional, normally distributed vector, e, with nonsingular variance-covariance matrix, V. The null hypothesis, H_0, is that the expected value of e is the zero vector, and the task is to define a unit normal variate which has the highest probability of detecting a true nonzero value. In the case of data reconciliation, the vector of residuals in the constraints is

$$e = B\tilde{x} \tag{1}$$

where the measurements, $\tilde{x}$, are used. The matrix, B, is the remaining constraint matrix for the measured species flow rates, after all of the unmeasured species and total flow rates have been removed, following Crowe (1986).

The alternative hypothesis, H_1, is that

$$E(e) = y \neq 0 \tag{2}$$

where y is the postulated gross error, normally a scalar times a unit vector. Define a general unit normal variate,

$$z_e(W, r) \triangleq \frac{r^T We}{\sqrt{r^T WVW^T r}} \tag{3}$$

which can be seen to have unit variance. The test, for any choice of W, is applied by calculating e and V from the measurements, $\tilde{x}$, the constraint matrix, B, and Σ, the variance matrix of the measurements. The vector, r, will normally be a unit vector for the test of a single contraint. If $|z_e(W, r)|$ exceeds the value of the standard normal variate at the selected confidence level, the null hypothesis, H_0, is rejected. The conventional constraint test has $W = I$. The essential result is contained in the following.

Lemma

Given hypothesis, H_1, with expected value of y and variance-covariance matrix, V, for e,

$$E[z_e(V^{-1}, y)] \geq \text{abs}\{E[z_e(W, y)]\} \quad (4)$$

for all square nonsingular W, and for all $r \neq y$,

$$E[z_e(V^{-1}, y)] \geq \text{abs}\{E[z_e(V^{-1}, r)]\} \quad (5)$$

Proof

The lemma may readily be proven using the Cauchy-Schwartz inequality in the form,

$$|v^T V^{-1} w| \leq \sqrt{v^T V^{-1} v}\sqrt{w^T V^{-1} w} \quad (6)$$

for nonzero m-vectors, v and w, and for positive definite matrix, V, with equality if and only if $v = \beta w$.

Remarks

• The test (MP) statistic, Eq. 3, with $W = V^{-1}$ has maximum power, since it will be expected to exceed, in absolute value, any other statistic for that error, as well as the MP statistic for any other error. Thus it has the greatest probability of rejecting the null hypothesis that there is no gross error, when there truly is one. Note however, that the inequalities on expected values do not mean that the analogous inequalities on actual values will always hold.

• Of course, an arbitrary choice of W would not lead to a physically meaningful linear combination of constraints. However, the optimal choice of the inverse of V, the variance matrix of e, does imply a test of the Lagrange multipliers, each directly associated with a constraint, since

$$\lambda = -V^{-1}e \quad (7)$$

This holds exactly for the linear case (Crowe et al., 1983) and approximately for the bilinear case (Crowe, 1986).

• While the MP constraint test is applicable, in theory, to arbitrary gross constraint errors, y, in practice, the test would be used to find a single gross error in a constraint in a specific set of constraints. The inequalities, Eqs., 4 and 5, are unchanged if y is multiplied by an arbitrary nonzero scalar. However, the power of a test statistic, Eq. 3, for any given W, is increased if that scalar exceeds unity.

• The value of the conventional statistic for the kth constraint,

$$z_{e,k} = z_e(I, u_k) = e_k/\sqrt{V_{kk}} \quad (8)$$

is invariant if the kth constraint is left unaltered, either during the deletion of a measurement or another constraint, or during the alteration of the other constraints by taking linear combinations of them with the kth constraint. On the other hand, the corresponding MP constraint statistic can change in such cases because the inverse of the modified matrix, V, differs, in general, from that of V itself, even when there is no covariance among measurements. This means that the MP constraint test statistic is, in general, a function of the *set* of constraints and is the reason the test should normally be applied with only y proportional to a unit vector. Evidence about gross error is, however, provided by the change in the MP test, caused by the deletion of a constraint.

• It is readily seen from Eq. 6 that the absolute maximum of the MP statistic is attained when $y = e$, since

$$\frac{y^T V^{-1} e}{\sqrt{y^T V^{-1} y}} \leq \sqrt{e^T V^{-1} e} \quad (9)$$

Thus, we are seeking a simple gross error, y, which accounts for as much as possible of the chi-square statistic, the square of the right side of Eq. 9.

Equality of *MP* Constraint and Measurement Tests

The MP measurement test statistic, as defined by Mah and Tamhane (1982) for the linear case, where $V = H$, is

$$z^*_{a,j} = \frac{B_j^T H^{-1} e}{\sqrt{B_j^T H^{-1} B_j}} \quad (10)$$

where B_j is the jth column of B. The MP constraint test statistic is

$$z^*_e(y) = \frac{y^T H^{-1} e}{\sqrt{y^T H^{-1} y}} \quad (11)$$

Clearly, the two MP statistics are equal, if

$$B_j = \beta y \quad (12)$$

for some nonzero scalar, β. Conversely, equality of the two statistics for any randomly measured e, implies Eq. 12, since we could collect m statistically independent samples of e, forming a nonsingular matrix on both sides of that equality. This matrix could then be eliminated from the equality. Thus, Eq. 12 is necessary and sufficient for the equality of an MP measurement statistic and an MP constraint statistic. This means that the MP measurement statistic for a feed or product stream, which corresponds to a $\pm$ unit vector for a column of B, would equal the MP constraint statistic for the one balance in which that stream appears, thus confounding the two tests.

Deletion of a Suspect Balance

If the MP statistic of a balance is too large, the balance would be suspected of having an unidentified flow or accumulation. The deletion of that balance is equivalent to adding an unmeasured flow to it. Crowe (1988) showed how the reconciliation after deletion of a measurement, could be found from the base case reconciliation. The same treatment applies here, and in particular, tells us that the reduction in the objective function from the deletion of a single balance is exactly equal to the square of its MP constraint statistic.

Examples

Two examples are presented, to compare the conventional and the MP constraint tests. In applying the test, we have the matrix, V, and its inverse from the reconciliation of the data, so

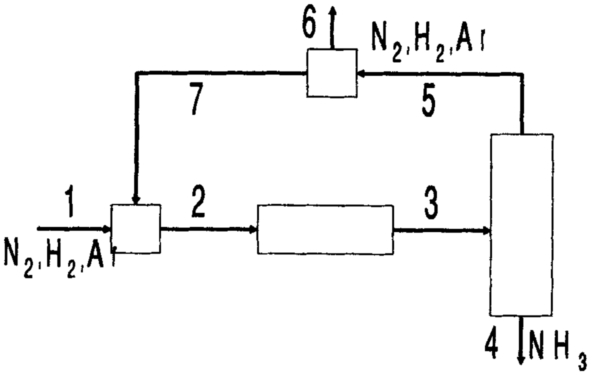

Figure 1. Ammonia synthesis loop, Example 1.

that the MP statistic is easy to compute. Example 1 is the ammonia synthesis loop of Crowe (1988), with the same measured data. Example 2 is an actual chemical extraction plant for which the reconciliation of total mass flow data was presented by Holly et al. (1989). The flow diagrams are given in Figures 1 and 2. All tests are assessed at the global 95% confidence level for jointly varying unit normal variates (Mah and Tamhane, 1982).

Example 1

The matrix, V, is identical to H for this case, and from the data in Crowe et al. (1983), is

$$V = \begin{bmatrix} 12.74 & -11.89 & 2.31 & -0.0163 \\ -11.89 & 12.64 & 1.14 & 0.0214 \\ 2.31 & 1.14 & 13.28 & 0.0142 \\ -0.0163 & 0.0214 & 0.0142 & 2.58E-4 \end{bmatrix}$$

Crowe (1988) showed that no single deletion of a measurement could lead to the passing of all statistical tests, but that three pairs of deletions did satisfy all tests. The constraint statistics are given in Table 1 for the original balances, I–IV, and for the modified balances, upon specific deletions of measurements. The conventional statistic is shown only once, since it is invariant for a given balance.

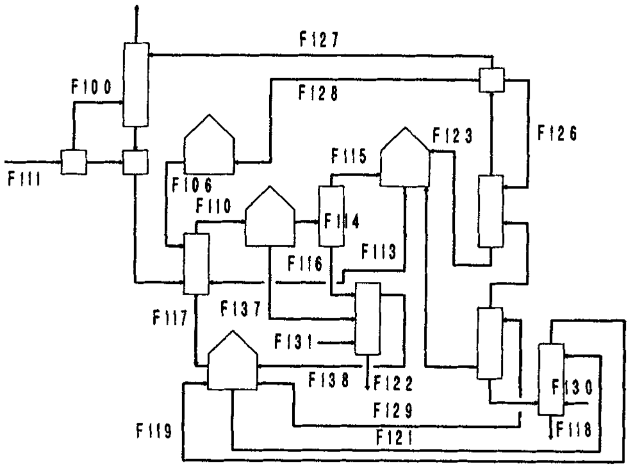

Figure 2. Chemical extraction plant, Example 2.

As seen in Table 1, the tests for constraints I and III are violated by the conventional statistic, but only the test for III is violated by the MP statistic. The MP tests are passed when only $H_2(1)$ is deleted. Among single deletions, this yielded the lowest chi-square value, which was, however, still statistically too large (Crowe, 1988), as were the MP measurement statistics for $N_2(2)$ and $N_2(3)$. Thus, the MP constraint test should be used in conjunction with the MP measurement test and the global chi-square test.

When $NH_3(4)$ alone is deleted, two MP tests are violated which are not under the conventional test, an example of additional information from the MP test. The MP constraint tests, as well as the measurement and chi-squre tests (Crowe, 1988), are all satisfied when each of the three most suspect pairs is deleted. The correct pair had a lower absolute maximum under each test than the other pairs. The fact that the correct and two incorrect pairs of deletions passed all of the tests, shows that ultimately,

Table 1. Constraint Statistics for Example 1, Ammonia Loop

Base Case	$N_2(1)$*	$H_2(1)$	$Ar(1)$	$N_2(2)$	$Ar(2)$	$N_2(3)$	$NH_3(4)$**	$H_2(5)$	
$B =$	0	0	0	1	0	−1	−0.5	0	(I)
	1	0	0	−1	0	0.98	0	0	(II)
	0	1	0	0	0	0	−1.5	−0.02	(III)
	0	0	1	0	−0.02	0	0	0	(IV)

Constraint	Delete None		Delete $H_2(1)$	Delete $N_2(1)$	Delete $NH_3(4)$	Delete $N_2(1)$, $NH_3(4)$	Delete $H_2(1)$, $N_2(3)$	Delete $H_2(1)$, $N_2(2)$
	$\|z_c\|$	$\|z_c^*\|$	$\|z_c^*\|$	$\|z_c^*\|$	$\|z_c^*\|$	$\|z_c^*\|$	$\|z_c^*\|$	$\|z_c^*\|$
I	2.62	1.53	2.25	1.94	—	—	—	—
II	1.83	2.08	1.25	—	3.43	—	—	—
III	4.60	4.15	—	3.80	—	—	—	—
IV	1.08	0.94	1.40	0.63	1.86	1.58	0.52	0.53
III-3*I	1.06	—	—	—	2.90	1.57	—	—
I+II	2.26	—	—	—	—	—	2.05	—
0.98* I+II	2.13	—	—	—	—	—	—	1.91

*+10% gross error.
**+20% gross error.

the unambiguous identification of gross errors requires the inspection of the measuring devices and procedures on the plant site.

Example 2

It is notable in Table 2 that all of the conventional constraint tests are satisfied and their values are strictly less than the corresponding *MP* statistics. By contrast, the first three *MP* tests with no deletion, are strongly violated. Single deletions of any of the three streams shown, led to satisfactory values of the *MP* tests. It was, however, noted by Holly et al. (1989) that the deletion of F131 led to its having to have a negative flow. Indeed, after deletion of either F115 or F138, all statistical tests were passed (Holly et al., 1989). Whether one of these measurements, or both, was actually in gross error when the data were taken, is not known.

Acknowledgment

The author is grateful to the Natural Sciences and Engineering Research Council of Canada for financial support. He also wishes to thank Professor R. G. H. Prince and the Department of Chemical Engineering, University of Sydney, for providing accommodation and facilities during the completion of this work.

Notation

B = matrix in constraint eq., mxn
e = $B\tilde{x}$, $mx1$
H = $B\Sigma B^T$, mxm
I = identity matrix
m = number of constraints
n = number of species flow rates
r, v, w = arbitrary vectors, $mx1$
u_k = unit vector, kth column of I
V = variance-covariance matrix of e, mxm
W = arbitrary square non-singular matrix, mxm
$\tilde{x}$ = vector of measured flow rates, $nx1$
y = expected value of e under alternative hypothesis, mx
z = unit normal variate

Greek letters

β = arbitrary scalar
λ = vector of Lagrange multipliers associated with constraints, $mx1$
Σ = variance-covariance matrix of measurements, nxn

Table 2. Constraint Statistics for Example 2

F111 − F118 − F122 = 0	(I)
F114 − F115 − F116 = 0	(II)
F116 − F122 + F131 + F137 − F138 = 0	(III)
F110 − F114 − F137 = 0	(IV)
F117 + F119 − F121 − F129 + F138 = 0	(V)
− F113 + F115 − F118 − F119 + F121 − F127 − F128 + F129 + F130 = 0	(VI)

			Delete		
Constraint	Delete None		F115	F131	F138
	$\lvert z_e\rvert$	$\lvert z_e^*\rvert$	$\lvert z_e^*\rvert$	$\lvert z_e^*\rvert$	$\lvert z_e^*\rvert$
I	1.72	3.53	2.20	1.70	2.00
II	1.57	5.02	—	1.47	2.17
III	0.14	4.83	1.72	—	—
IV	0.37	2.55	0.13	0.04	0.39
V	0.18	0.68	1.05	0.31	—
VI	0.40	0.88	—	0.57	1.30
II+VI	1.27	—	2.14	—	—
III+V	0.08	—	—	—	1.67

Subscripts

a = measurement test
e = constraint test

Superscripts

$*$ = maximum power (*MP*) test

Literature Cited

Crowe, C. M., Y. A. Garcia Campos, and A. Hrymak, "Reconciliation of Process Flow Rates by Matrix Projection: I. The Linear Case," *AIChE J.*, **29**, 881 (Dec., 1983).

Crowe, C. M., "Reconciliation of Process Flow Rates by Matrix Projection: II. The Nonlinear Case," *AIChE J.*, **32**, 616 (Apr., 1986).

———, "Recursive Identification of Gross Errors in Linear Data Reconciliation," *AIChE J.*, **34**, 541 (Apr., 1988).

Holly, W., R. Cook, and C. M. Crowe, "Reconciliation of Mass Flow Rate Measurements in a Chemical Extraction Plant," *Can. J. Chem. Eng.*, in press (1989).

Mah, R. S. H., and A. C. Tamhane, "Detection of Gross Errors in Process Data," *AIChE J.*, **28**, 828 (Sept., 1982).

Reilly, P. M., and R. E. Carpani, "Application of Statistical Theory of Adjustment to Material Balances," 13th Can. Chem. Eng. Conf., Montreal (1963).

Tamhane, A. C., "A Note on the Use of Residuals for Detecting an Outlier in Linear Regression," *Biometrika*, **69**, 488 (1982).

Manuscript received Nov. 21, 1988, and revision received Feb. 8, 1989.