Maximum Power Tests for Gross Error Detection Using Likelihood Ratios

Shankar Narasimhan
Department of Chemical Engineering
Indian Institute of Technology
Kanpur, India 208016

The problem of detecting and identifying gross errors in measurements due to biases in the measuring instruments and gross errors in steady state conservation constraints due to unknown leaks or unaccounted departures from steady state operation has been well studied. Several statistical tests for this purpose have been developed, such as, for example, the measurement test (MT) (Mah and Tamhane, 1982), the constraint test (CT) (Mah et al., 1976), and the generalized likelihood ratio (GLR) test (Narasimhan and Mah, 1987).

In order to achieve the best performance, it is important to apply the test that has the maximum power (the greatest probability of detecting the presence of a gross error when one is actually present) without increasing the probability of type I error (probability of wrongly detecting a gross error when none is present). Tamhane (1982) derived a maximum power (MP) measurement test for outlier detection in linear regression. The same test also gives the MP measurement test for detecting gross errors in measurements under the assumptions of

a. linear constraints
b. variables measured directly
c. all variables measured

Mah and Tamhane (1982) used the same argument to construct a test for the case when the variables are not directly measured, but did not show that this gives the MP measurement test. Recently, Crowe (1989) developed an MP constraint test using similar arguments, which is valid when the constraints are linear. The purpose of this note is to show that the MP measurement and MP constraint tests are equivalent to the GLR test for simple steady state models, and that the GLR test gives the MP measurement tests and MP constraint tests for original constraints even when the variables are not directly measured or when some variables are not measured.

Equivalence between GLR Test and MP Tests

The simple linear steady state model with no unmeasured variables and all variables directly measured is first considered. Following the arguments of Tamhane (1982), in order to show that a test gives the MP test statistic for identifying a single gross error when a gross error is present, it is sufficient to prove that the absolute expected value of the test statistic of the MP test for detecting that gross error is greater than or equal to the corresponding test statistic of any other test and is also greater than or equal to the test statistics for detecting other gross errors. In the subsequent proofs, instead of considering the GLR test statistic as proposed by Narasimhan and Mah (1987), we consider the square root of the GLR test statistic. This does not limit the generality of the results in any way.

Lemma 1. The GLR test is more powerful than an MT based on any singular or nonsingular linear transformation of the measurement residuals. In particular the MP measurement test is the same as the GLR test.

Proof. Let A be the constraint matrix and Q the covariance matrix of measurement errors and y be the vector of measurements. The measurement residuals are given by

$$a = QA'V^{-1}Ay \tag{1}$$

where

$$V = AQA' \tag{2}$$

and the constraint residuals are given by

$$r = Ay \tag{3}$$

Let the measurement residuals be linearly transformed using the matrix $\overline{A}$. Let a gross error of magnitude δ be present in measurement i. Then the expected value of the MT statistic for measurement i (which is used for detecting a gross error in measurement i) based on the transformed measurement residuals is given by

$$E[z_i] = \frac{\delta\, e_i'\overline{A}QA'V^{-1}Ae_i}{\sqrt{e_i'\overline{A}QA'V^{-1}AQ\overline{A}'e_i}} \tag{4}$$

while the expected value of the GLR test statistic, based on the constraint residuals, for measurement i is given by

$$E[T_i] = \delta \sqrt{e_i' A' V^{-1} A e_i} \tag{5}$$

By defining matrices

$$\begin{aligned} R'R &= V^{-1} \\ P &= RA \\ B &= RAQ\overline{A}' \end{aligned} \tag{6}$$

and using the Cauchy-Schwarz inequality

$$|v'w| \leq \sqrt{(v'v)(w'w)} \tag{7}$$

we can readily see that $E[T_i] \geq E[z_i]$. We can also easily show that the expected value of the test statistic for measurement i is greater than the expected value of the test statistic for any other measurement j. If $\overline{A} = Q^{-1}$, then the MP measurement test is obtained (Tamhane, 1982), and it can be readily seen from Eqs. 4 and 5 that, in this case, $E[z_i] = E[T_i]$. In fac. ʾboth the test statistics are equal, not just their expected values (Narasimhan and Mah, 1987).

Lemma 2. The GLR test is more powerful than a CT based on any singular or nonsingular linear transformation of the constraint residuals. In particular the MP constraint test is the same as the GLR test.

Proof. Let the constraint residuals be linearly transformed using the matrix $\overline{A}$. Let a gross error of magnitude δ be present in constraint i. Then the expected value of the CT for constraint i based on the transformed constraint residuals is given by

$$\begin{aligned} E[z_i] &= \frac{\delta\, e_i' \overline{A} e_i}{\sqrt{e_i' \overline{A} V \overline{A}' e_i}} \\ &= \frac{\delta\, e_i' \overline{A} V V^{-1} e_i}{\sqrt{e_i' \overline{A} V \overline{A}' e_i}} \end{aligned} \tag{8}$$

while the expected value of the GLR test statistic based on the constraint residuals for constraint i is given by

$$E[T_i] = \delta \sqrt{e_i' V^{-1} e_i} \tag{9}$$

Note that the GLR test statistic for a simple constraint test is obtained by using the vector e_i instead of the vector m_i for a leak as suggested by Narasimhan and Mah (1987).

By defining matrices

$$\begin{aligned} R'R &= V^{-1} \\ P &= (\overline{A} R^{-1})', \end{aligned} \tag{10}$$

and using the Cauchy-Schwarz inequality we can readily see that $E[T_i] \geq E[z_i]$. We can also easily show that the expected value of the test statistic for constraint i is greater than the expected value of the test statistic for any other constraint j. If $\overline{A} = V^{-1}$, then the MP constraint test is obtained (Crowe, 1989), and it can be readily seen from Eqs. 8 and 9 that, in this case, $E[z_i] = E[T_i]$. In fact, the test statistics of the two tests will be identical.

Thus the GLR test gives the most powerful test statistics for detecting gross errors in measurements or constraints. The use of deriving MP test statistics through the GLR test is as follows.

In obtaining the MP measurement test, Tamhane (1982) restricted the hypothesis to allow gross errors only in the measurements. Similarly, the MP constraint test derived by Crowe (1989) restricts the hypothesis to allow gross errors only in the constraints. However, if the user does not know a priori whether a gross error is present in the measurement or constraint, then the hypothesis has to be formulated to allow for both possibilities. In this case, the approach used by Tamhane (1982) and Crowe (1989) in deriving MP tests does not provide a basis for comparing the MP measurement with the MP constraint test statistics. On the other hand, since the GLR test hypothesis allows for gross errors in the measurements and constraints, the MP test statistics can be compared. In fact, the GLR test suggests that the maximum among the measurement and constraint test statistics should be chosen and an error in the corresponding measurement or constraint is detected if it exceeds a critical value, because this gives the maximum probability among all possible hypotheses of a single gross error.

MP Tests for General Steady State Models

Unmeasured variables

When all the variables are not measured, the constraints are given by

$$A_1 x + A_2 u = 0 \tag{11}$$

where x: $nx1$ is the vector of measured variables and u: $px1$ is the vector of unmeasured variables and A_2 is assumed to be of full column rank, p. As suggested by Crowe et al. (1983), the unmeasured variables can be eliminated by premultiplying the constraints by a projection matrix P: $(m\text{-}p)xm$ of rank $m\text{-}p$, where m is the number of constraints, to give

$$PA_1 x = Ax = 0 \tag{12}$$

If a gross error of magnitude δ is present in one of the original constraints i, then by using the same procedure as described in Narasimhan and Mah (1989), we can see that the expected value of the constraint residuals $= \delta P e_i$. If the vector $P e_i$ is not identically equal to 0, then the expected value of the GLR test statistic for detecting a gross error in the original constraint i is given by

$$E[T_i] = \delta \sqrt{e'_i P' (V)^{-1} P e_i} \tag{13}$$

Using arguments similar to those used in the proof of lemma 2, we can show that $E[T_i]$ is greater than or equal to the expected value of the CT statistic based on a singular or nonsingular linear transformation of the constraint residuals. Thus the GLR test gives MP test statistics for detecting gross errors in the original constraints even when unmeasured variables are present.

It should be noted that since there are fewer than m constraints in the projected constraint set, Eq. 12, the normal method of testing each constraint by using the CT gives a test for detecting a gross error in a projected constraint, but this does not indicate which of the original constraints contains the gross error. The procedure used by Crowe (1989) to derive MP constraint tests, in which all unmeasured variables are first eliminated by using a projection matrix, gives only MP test statistics for detecting a gross error in a projected constraint. However, Crowe (1989) has also shown that an MP constraint test is equivalent to a test of the corresponding Lagrangian multiplier. Therefore, if we retain the original constraints as such and solve for their corresponding Lagrangian multipliers, then a test of the Lagrangian multiplier gives an MP constraint test of the original constraints even when unmeasured variables are present (Crowe, personal communication).

Indirectly measured variables

When the variables are not measured directly, then by a simple redefinition of variables we obtain a problem with a larger number of constraints in which some variables are unmeasured but the others are all directly measured. If n is the number of measurements, p ($\leq n$) is the number of variables, and m ($\leq p$) is the number of constraints relating the variables, then following the procedure of Narasimhan and Mah (1989) we obtain a problem of $n + p$ variables of which n are directly measured and the remaining unmeasured, and $n + m$ constraints relating the variables. The proof in the previous section on unmeasured variables can be used to show that the GLR test statistics gives MP test statistics for detecting gross errors in measurements and original constraints even in this case.

It is worthwhile to compare the GLR test statistic with the test statistic used by Mah and Tamhane (1982) for the case when the measurement matrix is of full column rank. The authors used a different approach to compute the measurement residuals which were linearly transformed using Q^{-1}. The MT statistics were obtained based on transformed residuals. Using this method however, it is not easy to show that MP test statistics are obtained. Alternatively, the approach described in Narasimhan and Mah (1989) can be used to obtain the measurement residuals which is given by Eq. 1 where A now represents the projected constraint matrix. Using this form of the equation for the measurement residuals and lemma 1, it is now easy to prove that the MT statistics used by Mah and Tamhane (1982) are in fact the MP test statistics, which are equivalent to the GLR test statistics.

Acknowledgment

The author expresses his appreciation to C. M. Crowe for useful suggestions in reviewing this paper.

Notation

- a = measurement residuals
- A = constraint matrix, projected or otherwise
- $\bar{A}$ = matrix of a linear transformation
- A' = transpose of matrix A
- e_i = unit vector with unity in row i
- $E[a]$ = expectation of a
- m = number of constraints
- n = number of measurements
- p = number of variables
- P = projection matrix
- Q = convariance matrix of measurement errors
- r = constraint residuals
- T_i = GLR test statistic
- u = vector of unmeasured variables
- V = matrix, Eq. 2
- V^{-1} = inverse of matrix V
- x = vector of measured variables
- y = vector of measurements
- z_i = measurement or constraint test statistic
- 0 = vector of zeros

Greek letters

- δ = magnitude of gross error

Literature Cited

Crowe, C. M., Y. A. Garcia Campos, and A. Hrymak, "Reconciliation of Process Flow Rates by Matrix Projection. 1: The Linear Case," *AIChE J.*, **29**, 881 (1983).

Crowe, C. M., "Test of Maximum Power for Detection of Gross Errors in Process Constraints," *AIChE. J.*, **35**, 869 (1989).

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

Mah, R. S. H., G. M. Stanley, and D. M. Downing, "Reconciliation and Rectification of Process Flow and Inventory Data," *Ind. Eng. Chem. Process Des. & Dev.*, **15**, 175 (1976).

Narasimhan, S., and R. S. H. Mah, "Generalized Likelihood Ratio Method for Gross Error Identification," *AIChE J.*, **33**, 1514 (1987).

———, "Treatment of General Steady State Models in Gross Error Detection," *Comput. Chem. Eng.*, **13**, 851 (1989).

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

Manuscript received Nov. 11, 1989, and revision received July 12, 1990.