Reconciliation of Process Flow Rates by Matrix Projection

Part II: The Nonlinear Case

Flow rate and concentration measurements in a steady state process are reconciled by weighted least squares so that the conservation laws and other constraints are obeyed. Two projection matrices are constructed in turn, in order to decompose the problem into three subproblems to be solved in sequence. The first matrix eliminates all unmeasured component flow rates and concentrations from the equations; the second then removes the unmeasured total flow rates. The adjustments to component flow rates are iteratively determined, starting with guessed values of unmeasured total flow rates.

Chi-square and normal test statistics are derived by linearizing the equations, to allow detection of gross errors in imbalances and adjustments of measurements.

C. M. CROWE

Department of Chemical Engineering
McMaster University
Hamilton, Ontario, Canada L8S 4L7

SCOPE

The cornerstone for monitoring plant performance is a set of steady state balances for component and total flow rates. Such flow rates are normally obtained from measurements of total flow rates and of concentrations, which are subject to random and sometimes gross errors, and thus in general violate conservation laws. The measurements should be reconciled, in some "best" sense, to obey those laws and any other constraints that are required to be enforced.

The linear case, where it is assumed that the total flow rate is measured in any stream in which a concentration is measured, was discussed by Crowe et al., (1983). In this paper, that assumption is omitted so that the balance equations contain products of unknowns and thus are nonlinear (actually bilinear).

Conflict with conservation laws can only arise in sections of the process where measurements have been made in all streams in a balance equation. Thus previous work by Vaclavek (1969), Mah et al. (1976), and Romagnoli and Stephanopoulos (1981) was directed to devising path-tracing or combinatorial algorithms for efficiently finding a maximal set of sections of the process with redundant measurements—the reduced balance scheme (RBS). Unfortunately, such searches must be done separately for each component, or each element in the case of reactions. Crowe et al. (1983) proposed a constructive method of finding the RBS by a projection matrix that effectively blanked out all unmeasured flows in the linear case.

The approach here is to extend the technique of Crowe et al. to the nonlinear case by constructing two successive projection matrices. The first eliminates all unmeasured component flow rates and concentrations; the second then eliminates the unmeasured total flow rates from the balance equations. The original problem is thus divided into three sequentially solved subproblems. Guesses of the unmeasured total flow rates are used to solve for the adjustments to component flow rates. These guesses are iteratively updated in the second subproblem until convergence is achieved. Then those unmeasured component flow rates that are determinable are found in the third subproblem.

Gross errors in the measurements can arise from instrument malfunction or miscalibration, sampling errors, and unsuspected leaks or departures from steady state. Measurements that are in gross error should be detected before the data are reconciled so that they can be corrected or eliminated. Since the equations are bilinear, the statistical analysis requires local linearization about the measurement values, in order to find test criteria for the imbalances in component flow rates and for the adjustments to particular measured values.

CONCLUSIONS AND SIGNIFICANCE

A pair of projection matrices has been applied to the equations defining the conservation laws and other constraints. This approach finds a maximal set of balances on which reconciliation can be done. There are no restrictions imposed upon the placement of measurements, nor upon what is measured in any one stream. The construction of these projection matrices is direct and each requires the inversion of a matrix that is also needed for subsequent steps in the solution.

Statistical tests for the set of imbalances or residuals in the constraints (chi-square) and for individual imbalances and ad-

justments of measurements (unit normal variate) are developed by linearizing the constraint equations. These tests can then be used jointly to find which imbalances or adjustments are grossly in error.

INTRODUCTION

The problem to be addressed is that of reconciling measurements of total flow rates and concentrations such that the conservation laws and possibly other constraints are obeyed, by a "best" choice of adjustments to the measurements. In an earlier paper, Part I, Crowe et al. (1983) presented a solution to the linear case where it was assumed that the total flow rate in any stream was measured if any concentrations were measured in that stream. A summary of the published work done on the linear case was given in Part I, and more general reviews have been presented by Hlavacek (1977) and Mah (1981).

In this paper, no restrictions are imposed upon the location of streams in which total flow rates or concentrations are measured nor upon which concentrations are measured in any stream. As a result, the balance equations will, in general, contain products of pairs of unknowns and thus will be bilinear. Vaclavek et al. (1976a,b) developed the theoretical basis of the bilinear problem and discussed conditions under which the original balance equations could be reduced in number. They assumed that either all or none of the concentrations in a stream were measured and thus could partition the streams into four categories, namely:

Category	Total Flow	Concentrations
1	M	M
2	U	M
3	M	U
4	U	U

where M = measured, U = unmeasured.

However, with an arbitrary distribution of measurements, the classification must refer to components in streams, in which case categories 3 and 4 can be combined since the component flow rate is unmeasured in each case. Thus, there will be three categories of variables:

1. Total flow rate and concentration measured and adjustable.
2. Concentration measured and adjustable.
3. Total flow rate unknown or measured; component flow rate unmeasured.

The classification by Vaclavek et al. (1976a) of unknown variables into determinable and indeterminate can be extended here in that each of categories 2 and 3 separately may be so classified. The indeterminate variables in each case will correspond to dependent columns of the respective matrices in the balance equations. Vaclavek et al. also divided the measurements according to whether they should be corrected or not. This division will depend not only on the flow sheet but on the variance-covariance (dispersion) matrix of the measurements. Only if the dispersion matrix is diagonal can this division be done from the flow sheet alone.

In order to reduce the number of balance equations to a minimum number, without sacrificing information, Vaclavek et al. proposed a two-step reduction. The formulation presented here will lead to a different two-step reduction scheme which first eliminates variables of category 3 and other unmeasured variables, such as extents of reaction, and then eliminates the total flow rates corresponding to variables of category 2.

Stanley and Mah (1981a) developed the concepts of global and local observability of state variables, given a set of measurements and constraints, for the nonlinear problem. They also (1981b) applied graph theory to mass-energy flow networks to classify unmeasured variables as globally (or locally) observable or unobservable. The global (or local) redundancy of a measurement was determined by the effect of its deletion on the global (or local) observability.

Knepper and Gorman (1980) presented the theory of nonlinear estimation of constrained data sets and provided statistical tests for the estimate's being free from gross error. Romagnoli (1983) recently discussed a recursive technique for sequential addition or deletion of either measurements or constraints.

Beckmann (1982) used nonlinear programming to adjust measurement data in batch and steady state processes. Data reconciliation of an operating chemical process was studied by Ham et al. (1979), and of mineral and metallurgical processes by Cutting (1976), Hodouin and Everell (1980), Hodouin et al. (1982), Mular et al. (1976), Ragot and Aubrun (1980), Smith and Ichiyen (1973), White et al. (1977) and Wiegel (1972). A computer program for nonlinear least squares estimation was reported by Laguitton (1980).

PROBLEM STATEMENT

If vector x represents the n true component and total flow rates in all streams of a process, the m material balances and other constraints among them can be written as

$$Bx + S^T\xi = 0 \tag{1}$$

where matrix S represents the $(p \times m)$ stoichiometric matrix and ξ is the composite vector of p extents of reaction for all reactions in the process, as described in Part I.

The columns of the $(m \times n)$ matrix B are then permuted and partitioned so that

$$B \rightarrow [B_0|B_1|B_2|B_3]$$

where columns of B_0 correspond to exactly known flow rates k and those of B_i ($i = 1, 2, 3$) to n_i components in categories 1, 2, and 3 respectively.

The flow of component c in stream j, in category 1 is defined by

$$x_{cj} = M_j c_{cj} \tag{2}$$

where M_j is total flow rate and c_{cj} is concentration.

The measured values will be represented by $\tilde{}$ over the symbol and the estimated values by $\hat{}$. Thus,

$$\hat{x}_{cj} = \tilde{x}_{cj} + a_{cj} \tag{3}$$

where a_{cj} is the adjustment in stream j for the measured flow rate of component c.

For a stream with unmeasured total flow rate, we use d, δ for the measured concentration and its adjustment. Then,

$$\hat{d}_{c\ell} = \tilde{d}_{c\ell} + \delta_{c\ell} \tag{4}$$

for the estimated concentration of component c in stream ℓ. The unknown total flow rate will be N_ℓ.

As in Part I, we define

$$P = [B_3|S^T] \tag{5}$$

and represent all unknown component flow rates and extents of reaction by vector v. Then the constraints must be obeyed by the estimated values, so that

$$B_0k + B_1(\tilde{x} + a) + B_2N(\tilde{d} + \delta) + Pv = 0 \tag{6}$$

where N is the diagonal matrix of unknown flow rates, with as many entries for each stream as there are measured concentrations in that stream. The reconciliation problem can then be defined.

Problem P1

$$\min_{a,\delta,N,v} F(a,\delta) \triangleq a^T\Sigma_1^{-1}a + \delta^T\Sigma_d^{-1}\delta \tag{7}$$

subject to Eq. 6. Here, Σ_1 and Σ_d are estimates of the respective variance-covariance matrices.

There are several steps that can be taken to simplify this problem. The first is to define, as in Part I, the matrix Y such that

$$P^TY = 0 \tag{8}$$

with the columns of Y forming a basis for the null space of P^T. Then, Eq. 6 can be reduced to

$$Y^T[B_0k + B_1(\tilde{x} + a) + B_2N(\tilde{d} + \delta)] = 0 \tag{9}$$

This is the first stage of the reduction in the number of equations by eliminating all category 3 variables. The construction of Y can be done, as in Part I, by partitioning

$$P = \begin{bmatrix} P_1 & P_3 \\ P_2 & P_4 \end{bmatrix} \tag{10}$$

where P_1 is a maximal square nonsingular matrix contained in P. We can then choose

$$Y^T = [-P_2P_1^{-1}|I] \tag{11}$$

By definition, the righthand column partition of P depends on the lefthand columns, so that Eq. 8 is satisfied. The dependent columns of P correspond to indeterminate variables in category 3.

It should be noted that when P_1 is not in the upper left corner, the columns of each partition of Y^T in Eq. 11 should be permuted to correspond respectively to the rows of P_1 and P_2.

The second step of simplication involves defining a matrix Z that is a basis for the null space of

$$D^T \triangleq [B_{21}\tilde{d}_1 B_{22}\tilde{d}_2 \ . \ . \ . \ B_{2\ell}\tilde{d}_\ell \ . \ . \ . \ ]^TY \tag{12}$$

where $B_{2\ell}$ is the set of columns of B_2 corresponding to stream ℓ, and d_ℓ is the vector of measured concentrations in that stream. Then,

$$D^TZ \triangleq 0 \tag{13}$$

and

$$Dn = Y^TB_2N\tilde{d} \tag{14}$$

where n is the vector of distinct unknown total flow rates in category 2, so that Eq. 9 can be simplified to

$$Z^TY^T[B_0k + B_1(\tilde{x} + a) + B_2N\delta] = 0 \tag{15}$$

This then further reduces the size of the problem by eliminating total flow rates in category 2. We can represent Z analogously to Y, namely if

$$D = \begin{bmatrix} D_1 & D_3 \\ D_2 & D_4 \end{bmatrix} \tag{16}$$

with D_1 a maximal square nonsingular submatrix in D.

$$Z^T = [-D_2D_1^{-1}|I] \tag{17}$$

The total flow rates that correspond to dependent columns of D are indeterminate.

There are several advantages to defining Y and Z separately. First, the inverses of P_1 and D_1 are more efficient to compute separately than that of a larger combined matrix. Secondly, separate conditions are obtained for indeterminacy of variables in category 3 and of total flow rates in category 2. Thirdly, Y is a constant matrix so that Z contains all of the statistical variability due to that of $\tilde{d}$.

The third step in the simplification involves rewriting the second term in the objective function, Eq. 7, as

$$\delta^T\Sigma_d^{-1}\delta = (N\delta)^T(N\Sigma_dN)^{-1}(N\delta) \tag{18}$$

If estimates are made of N, say N_0, then define

$$\Sigma_2 \triangleq N_0\Sigma_dN_0 \tag{19}$$

so that problem P1 can be restated as P2.

Problem P2

(a) Solve:

$$\min_{a,(N\delta)} G(a,N\delta) \triangleq a^T\Sigma_1^{-1}a + (N\delta)^T\Sigma_2^{-1}(N\delta) \tag{20}$$

subject to Eq. 15.

(b) From the solution to part (a), solve Eq. 9, rewritten with Eq. 14 as

$$Dn = -Y^T[B_0k + B_1(\tilde{x} + a) + B_2(N\delta)] \tag{21}$$

for n.

(c) From the solutions to parts (a) and (b), solve

$$Pv = -[B_0k + B_1(\tilde{x} + a) + B_2N(\tilde{d} + \delta)] \tag{22}$$

for v.

The necessary condition for a solution to exist for each of Eqs. 21 and 22 is satisfied by the definitions of Y and Z. If the indeterminate variables are set to zero, the remaining unknowns can be solved for by

$$n_d = D_1^{-1}[I|0]R(21) \tag{23}$$

and

$$v_d = P_1^{-1}[I|0]R(22) \tag{24}$$

where $R(21)$, $R(22)$ are the righthand sides of Eqs. 21 and 22, of which we retain only the rows corresponding to D_1, P_1, respectively.

Since we already have the inverses of D_1 and P_1, the equations are efficiently solved. The space of all solutions to Eqs. 21 and 22 can be found from the fact that

$$D\left[\frac{-D_1^{-1}D_3}{I}\right] = 0 \tag{25}$$

Thus

$$n = \left[\frac{n_d}{0}\right] + \left[\frac{-D_1^{-1}D_3}{I}\right]z \tag{26}$$

for arbitrary z. Similarly, a general solution for v can be written. There is no guarantee that the solutions for n_d or v_d will be nonnegative so that the general solutions could be used to find vectors n and v, which conform to additional specifications such as nonnegativity.

Problem Solutions

It remains for us to solve problem P2(a). The Lagrangian L_2 is defined by

$$L_2 \triangleq \frac{1}{2}[a^T\Sigma_1^{-1}a + (N\delta)^T\Sigma_2^{-1}(N\delta)] - \\ -\lambda^TZ^TY^T[B_0k + B_1(\tilde{x} + a) + B_2(N\delta)] \tag{27}$$

with Lagrange multiplier vector, λ.

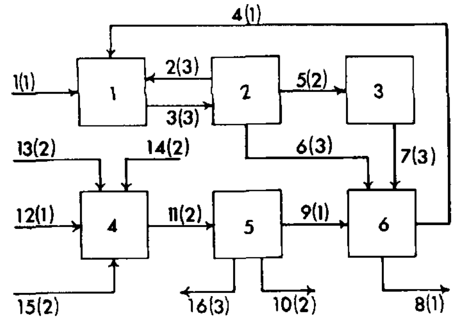

Figure 1. Flow diagram of Vaclavek's example s(j): Stream s(Category j).

Then derivatives are taken with respect to a and $(N\delta)$ and set equal to zero, giving respectively,

$$a = \Sigma_1 B_1^T Y Z \lambda \tag{28}$$

and

$$N\delta = \Sigma_2 B_2^T Y Z \lambda \tag{29}$$

Note that if D has full row rank, $Z = 0$, $a = 0$, and $N\delta = 0$. From Eq. 15,

$$Z^T Y^T [B_1 a + B_2 N\delta] = -Z^T Y^T [B_0 k + B_1 x] \tag{30}$$

From Eqs. 28 and 29, with

$$H \triangleq Y^T (B_1 \Sigma_1 B_1^T + B_2 \Sigma_2 B_2^T) Y \tag{31}$$

we obtain

$$\lambda = -(Z^T H Z)^{-1} Z^T Y^T [B_0 k + B_1 \tilde{x}] \tag{32}$$

This allows computation of λ, a, and $N\delta$ from Eqs. 32, 28, and 29. Problem P2(b) can then be solved for n. Now, because Σ_2 was computed using a guess for N (or n), we can iterate by updating Σ_2 and solving P2(a) again. This shows a fourth advantage of splitting matrices Z and Y, namely that this iteration does not involve variables of category 3.

If we wish to apportion the adjustments a_j of component flow rates x_j in stream j among the total flow rate M_j and the concentrations c_j, we will require

$$\hat{c}_j = \hat{x}_j / \hat{M}_j \tag{33}$$

This then leaves only one further condition to be specified. Having suppressed the subscript j for simplicity, we can list the following possible conditions:

(a) A particular concentration c_i is exactly known. Then

$$M = x_i / c_i \tag{34}$$

and Eq. 33 gives the other concentrations.

(b) Total flow rate M is exactly known. Then c are all calculated from Eq. 33.

(c) The concentrations must add up to a specified quantity ρ; for example, the mass or mole fractions of all species in the stream. Thus

$$1^T \hat{c} = \rho \tag{35}$$

where 1 is a vector of ones. Then

$$M = (1^T \hat{x}) / \rho \tag{36}$$

and c are found from Eq. 33.

(d) In the absence of any of the above conditions, we can minimize the weighted sum of squares of the adjustments as follows:

$$\underset{M}{\text{Min}}\, J = [(\hat{c} - \tilde{c})^T \Sigma_c^{-1} (\hat{c} - \tilde{c}) + (\hat{M} - \tilde{M})^2 / \sigma_m^2] \tag{37}$$

Note that J is a function only of M because c can be eliminated by Eq. 33.

If we set $dJ/dM = 0$, we obtain

$$\hat{M}^4 - \tilde{M}\hat{M}^3 + \sigma_M^2 (\hat{M}\tilde{c} - \hat{x})^T \Sigma_c^{-1} \hat{x} = 0 \tag{38}$$

which can be readily solved using Newton's method, with an initial guess of $\hat{M} = \tilde{M}$.

APPLICATION TO VACLAVEK'S EXAMPLE

The flow scheme of the example of Vaclavek et al. (1976b) is shown in Figure 1. In this case, all concentrations were assumed to be measured or none were, so that the streams themselves can be classified according to categories 1–3. The category of each stream is shown in parentheses. The matrices B_1, B_2 and B_3 are shown in Table 1, with I as the identity matrix of order equal to the number of components and the remaining entries zeroes. The matrix Y^T can then be defined as:

$$Y^T = \begin{bmatrix} \cdot & \cdot & \cdot & I & \cdot & \cdot \\ I & I & I & \cdot & \cdot & I \end{bmatrix} \tag{39}$$

We note that the remaining balances involve node 4 alone and nodes 1, 2, 3, and 6 combined.

Now, the matrix D can be written as

$$D = \begin{bmatrix} -\tilde{d}_{11} & \tilde{d}_{13} & \tilde{d}_{14} & \tilde{d}_{15} \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{40}$$

In Vaclavek's case, there were four components so that D had full column rank. In that case,

$$Z^T = [0_4 | I_4] \tag{41}$$

Thus, adjustments would be zero for the streams incident to node 4 and adjustments to flows in streams 1, 8, and 9 could be found as problem P2(a). Then estimates of the total flows of streams of category 2 incident to node 4 can be found.

Now, if there were fewer than four components, one or more total flows in streams 11, 13, 14, and 15 would be indeterminate but the remainder could be estimated. If there were more than four components, the null matrix in Eq. 41 would be replaced by $-D_2 D_1^{-1}$ as discussed above. Then, some balance envelope(s) around node 4 would be retained in the reduced

TABLE 1. MATRICES B_1, B_2, B_3 FOR VACLAVEK'S EXAMPLE

Node	B_1: 1	4	8	9	12	B_2: 5	10	11	13	14	15	B_3: 2	3	6	7	16
1	I	I	.	.	.	.	.	.	.	.	.	I	$-I$	.	.	.
2	.				.	$-I$	.	.	.	.	.	$-I$	I	$-I$	.	.
3	.				.	I	.	.	.	.	.	.	.	.	$-I$	.
4	.	.	.	.	I	.	.	$-I$	I	I	I	.	.	.	.	.
5	.	.	.	$-I$	.	.	$-I$	I	.	.	.	.	.	.	.	$-I$
6	.	$-I$	$-I$	I	.	.	.	.	.	.	.	.	.	I	I	.

problem P2(a), adjustments to the streams incident to it would not all be zero, and the total flow rates could again be estimated.

By comparison, Vaclavek et al. excluded node 4 from the reduced balance scheme at the first stage, where node 1 was retained alone, and nodes 2, 3 and 6 were combined as balance envelopes. On the second stage of reduction, these four nodes were combined. Vaclavek et al. imposed the condition that the sum of concentrations in any stream equal unity. This would not affect the results if the measured concentrations were normalized in advance. However, without such a priori normalization, an additional constraint for each stream in category 3 would be added, which in turn, would add a row to Z^T. Then the adjustments to concentrations in streams 11, 13, 14, and 15 would be nonzero, as expected.

STATISTICAL TESTS FOR GROSS ERRORS

The statistical analysis is complicated by the dependence of Z on the random vector d. In the Appendix, a Taylor series analysis is used to derive approximate normal distributions of

$$e \triangleq Z^T Y^T B_1 \tilde{x} \sim N(0,T) \quad (42)$$

$$a \sim N(0,Q_1) \quad (43)$$

and

$$\delta \sim N(0,Q_2) \quad (44)$$

with Q_1, Q_2 usually singular.

The data can then be tested as a whole since

$$\chi_e^2 \triangleq e^T T^{-1} e \quad (45)$$

is a chi-square statistic with the number of degrees of freedom equal to the rank of Z.

As in Part I, linear combinations of e can be tested against the unit normal distribution since

$$z \triangleq \frac{w^T e}{(w^T T w)^{1/2}} \sim N(0,1) \quad (46)$$

In particular, for $w = u_k$, the kth unit vector

$$z_k^e = \frac{e_k}{T_{kk}^{1/2}} \quad (47)$$

Individual elements of a, δ can be tested analogously by

$$z_j^a \triangleq \frac{a_j}{(Q_{1,jj})^{1/2}} \sim N(0,1) \quad (48)$$

and

$$z_\ell^\delta \triangleq \frac{\delta_\ell}{(Q_{2,\ell\ell})^{1/2}} \sim N(0,1) \quad (49)$$

By comparing elements of e, a, and δ whose absolute values are too large to be consistent with Eqs. 46, 48, and 49, one has information with which to diagnose which measurements are the most likely cause of the apparent gross errors.

One disadvantage of using z_k^e as a test statistic for the residuals in the constraints is that Z^T is not simply related to balances around nodes of individual species. An alternative is to define (with $k = 0$)

$$\epsilon \triangleq Y^T[B_1 \tilde{x} + B_2 N \tilde{d}] \quad (50)$$

By using Eqs. 9, 28, and 29, we can obtain

$$\epsilon = -HZ\lambda \quad (51)$$

and with Eqs. 32 and 42

$$\epsilon = HZT^{-1}e \quad (52)$$

Thus ϵ has a degenerate normal distribution,

$$\epsilon \sim N(0, HZT^{-1}Z^T H), \quad (53)$$

whose elements can be tested by

$$z_k^\epsilon \triangleq \epsilon_k/(HZT^{-1}Z^T H)_{kk}^{1/2} \quad (54)$$

Mah and Tamhane (1982) have discussed the power of such tests. They pointed out that to require each e_k, a_j, and δ_ℓ to lie within a given level of confidence would lead to unwarranted rejection of a value which is consistent with $N(0,1)$, an error of type I. They suggested, following Sidak (1967), setting a confidence level β for each measurement of a set of s independent measurements such that jointly, the confidence level was α, with

$$\beta \triangleq 1 - (1 - \alpha)^{1/s} \quad (55)$$

Sidak showed that this would provide a test for the α confidence level of the distribution of the maximum of such a set, which gives an upper bound on a type I error.

Then in testing elements of e using z_k^ϵ, and in the light of Eq. 42, one would find the value of β using $s = r_z$, the rank of Z (and T).

However, for tests of z_j^a, z_ℓ^δ, and z_k^ϵ, not only are the variance matrices singular with rank r_z, but a and δ are jointly dependent on ϵ and on e, through Eqs. A16 and A17, which give

$$-Z^T Y^T (B_1 a + B_2 N \delta) = Z^T \epsilon = e \quad (56)$$

It seems reasonable to suggest again using $s = r_z$ to find β for testing z_j^a, z_ℓ^δ, and z_k^ϵ, although Iordache et al. (1984) have argued for a value intermediate between r_z and the total number of elements in a and δ or in ϵ, respectively. In any event, the practical aim is to flag suspect values of e, a, δ, and ϵ so that they can be reexamined and any faults in the measurements can be identified. When the tests with one set of measurements do not clearly identify suspect values, tests using averages of several sets of successive measurements should help to reveal any persistent gross errors since the variances are all reduced in proportion to the number of such sets.

DISCUSSION

The main advantage of the proposed method is that it reduces the computing load required for matrix inversion. The determination of Y and Z need only be done once and only the inversion of $(Z^T HZ)$ is required at each iteration to solve for total flow rates n.

The formulation as problem P2 has the advantage over P1 of separating the calculations of a and $(N\delta)$ from that of N itself. In problem P1, they must be solved together and Z, if used, will change at each iteration.

The savings in matrix inversion are substantial compared to the general method of Britt and Luecke (1973), which would require the inversion of

$$(B_1 \Sigma_1 B_1^T + B_2 \Sigma_2 B_2^T)$$

at each iteration, instead of $(Z^T HZ)$.

A computer program to implement the method was written and tested on a VAX 11/780 computer. Among the literature cases studied were those of Cutting (1976), Ham et al. (1979), Mular et al. (1976), Smith and Ichiyen (1973), and Wiegel (1972). The results corresponded reasonably closely to those reported by these authors, when direct comparison was possible.

The flotation circuit of Smith and Ichiyen (1973), shown here in Figure 2, is a useful example to illustrate the application of theory. Concentrations of Cu and Zn were measured in all streams except number 8 but no total flow rates are given, so

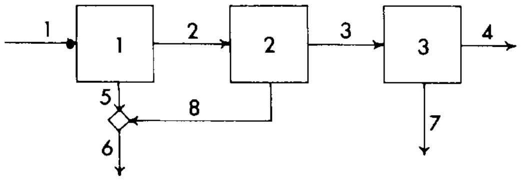

Figure 2. Smith and Ichiyen's flotation circuit Unit 1, unit cells; Unit 2, Cu flotation; Unit 3, Zn flotation.

that only relative flows can be estimated. Thus we define the feed flow to be unit mass. The B matrices are then

Stream: 1 2 3 4 5 6 7

$$B = \begin{bmatrix} I \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}; \quad B_2 = \begin{bmatrix} -I & \cdot & \cdot & -I & \cdot & \cdot \\ I & -I & \cdot & \cdot & \cdot & \cdot \\ \cdot & I & -I & \cdot & \cdot & -I \\ \cdot & \cdot & \cdot & I & -I & \cdot \end{bmatrix},$$

8

$$B_3 = \begin{bmatrix} \cdot \\ -I \\ \cdot \\ I \end{bmatrix}$$

where I is the (3×3) identity matrix so that three balance equations are imposed for each of the four nodes (i.e., Cu, Zn and total mass).

The matrix Y can then be seen to be

$$Y^T = \begin{bmatrix} I & \cdot & \cdot & \cdot \\ \cdot & I & \cdot & I \\ \cdot & \cdot & I & \cdot \end{bmatrix}$$

Matrix D is given by

$$D = \begin{bmatrix} -d_2 & \cdot & \cdot & -d_5 & \cdot & \cdot \\ d_2 & -d_3 & \cdot & d_5 & -d_6 & \cdot \\ \cdot & d_3 & -d_4 & \cdot & \cdot & -d_7 \end{bmatrix}$$

where $d_i = [d_{Cu} \quad d_{Zn} \quad 1]_i^T$ for stream i, in mass fractions. The relative standard deviation of the measurements was the same as in Smith and Ichiyen (1973), namely 6.56%.

The results of the reconciliations are shown in Table 2. In case 1, all the data were used and it is clear that several values of z^a are too large, as is the chi-square value. Even more disturbing is the fact that the Zn concentration in stream 8 must be negative. This was also true in Smith and Ichiyen's results although they do not remark on it.

In case 2, the concentration of Zn in stream 1 is assumed to be unknown. Although this results in acceptable values of the test statistics, the negative concentration of Zn remains.

In case 3, we imposed a further restriction that $n_{Zn,8} = 0.001 n_{Zn,2}$ in order to prevent a negative concentration. The result is that concentrations in streams 5 and 6 and $d_{Cu,1}$ now have values of z^a that are too high. It is clear that there is something wrong with the data, and indeed Smith and Ichiyen noted that the X-ray analyses for Zn were in error. There is in fact a second set of data taken six minutes earlier, given in Figure 2 of Smith and Lewis (1969). If one repeats the adjustments with the average of these two sets, the faultiness of the data is even more glaring, since the variances are all halved.

CONCLUSION

The theoretical approach to solving the problem of reconciling flow measurement data for the linear case, as presented in Part I, has been extended to the nonlinear case. The unknown component flow rates and extents of reaction are deleted, as in the linear case, by the constant projection matrix Y. Then, the unknown total flow rates are deleted by a second projection matrix, Z, which is stochastic because of the definition, Eq. 13, in terms of measured concentrations, d. The original reconciliation problem is then reformulated as a sequence of three subproblems, P2(a), (b), and (c). The first two are iteratively solved in alternation, and after convergence the third is solved separately.

ACKNOWLEDGMENT

The author is grateful to the Natural Sciences and Engineering Research Council, Ottawa, Canada, for financial support. He wishes to thank P. Rosza, Technical University of Budapest, Hungary, for his interest in this work and for many stimulating discussions. He also wishes to acknowledge the assistance of C. Feierabend, McMaster University, in preparing the computer program.

TABLE 2. SMITH AND ICHIYEN (1973) FLOTATION CIRCUIT

		Stream 1	2	3	4	5	6	7	8
	$\tilde{d}_{Cu}$%	1.93	0.45	0.13	0.09	19.86	21.44	0.51	—
	$\tilde{d}_{Zn}$%	3.81	4.72	5.36	0.41	7.09	4.95	52.10	—
Case 1	d_{Cu}%	1.928	0.45	0.128	0.09	19.88	21.43	0.513	35.36*
(all data)									
$\chi^2 = 13.1$	$\lvert Z_{Cu}\rvert$	0.38	0.77	0.38	0.38	0.77	2.01	0.38	—
$(\chi^2_{3,0.95} = 7.8)$	d_{Zn}%	4.95	4.78	4.95	0.41	7.07	4.92	51.88	−14.4*
	$\lvert Z_{Zn}\rvert$	3.55	0.77	1.91	0.38	0.77	2.01	0.38	—
	Total Flow	1.0	0.924	0.915	0.835	0.0761	0.0845	0.0808	0.00844
Case 2	d_{Cu}%	1.933	0.45	0.13	0.09	19.84	21.44	0.511	36.36*
($d_{Zn,1}$ unknown)	$\lvert Z_{Cu}\rvert$	0.72	0.72	0.11	0.11	0.72	0.72	0.11	—
$\chi^2 = 0.52$	d_{Zn}%	5.20*	5.04	5.22	0.41	7.11	4.94	52.03	−15.4*
	$\lvert Z_{Zn}\rvert$	—	0.72	0.68	0.11	0.72	0.72	0.11	—
$(\chi^2_{2,0.95} = 6.0)$	Total Flow	1.0	0.923	0.915	0.830	0.0765	0.0846	0.0854	0.00817
Case 3	d_{Cu}%	1.86	0.46	0.129	0.09	21.2	20.13	0.51	16.3*
($d_{Zn,1}$ unknown)	$\lvert Z_{Cu}\rvert$	2.57	2.67	0.19	0.092	2.67	2.71	0.09	—
$n_{Zn,8} = 0.001^* n_{Zn,2}$	d_{Zn}%	5.18*	5.08	5.18	0.41	6.64	5.23	52.04	0.25
$\chi^2 = 7.83$	$\lvert Z_{Zn}\rvert$	—	0.93	0.92	0.09	2.57	2.57	0.09	—
$(\chi^2_{3,0.95} = 7.8)$	Total Flow	1.0	0.932	0.913	0.829	0.0677	0.0867	0.0843	0.019

* Value required to satisfy balance.

NOTATION

a = vector of adjustments to measured component flow rates $(n_1 \times 1)$
B = balance matrix of process $(m \times n)$
B_0 = columns of B corresponding to fixed component flow rates $(m \times n_0)$
B_1 = columns of B corresponding to measured component flow rates $(m \times n_1)$
B_2 = columns of B corresponding to components with measured concentrations in streams with unmeasured total flow rates $(m \times n_2)$
B_3 = columns of B corresponding to components with unmeasured concentrations $(m \times n_3)$
c = concentration vector corresponding to $B_1(n_1 \times 1)$
d = measured concentrations corresponding to $B_2(n_2 \times 1)$
D = matrix, Eq. 12 $(r_Y \times s_2)$
E = expectation operator
e = imbalance vector, Eq. A1
H = matrix, Eq. 31
I = identity matrix
J = objective function, Eq. 37
k = vector of fixed component flow rates, corresponding to $B_0(n_0 \times 1)$
L = Lagrangian, Eq. 27
m = number of constraints
M = total flow rates of streams with measured component flow rates $(s_1 \times 1)$
n_i = number of variables in category $i = 0, 1, 2, 3$
n = vector of unmeasured flow rates of streams with measured concentrations, $d(s_2 \times 1)$
N = diagonal matrix, with each component of n repeated to correspond to the number of measured concentrations in that stream $(n_2 \times n_2)$
p = number of independent reactions
$P = [B_3 : S^T]$
Q_i = matrices, Eqs. A20 and A22 for $i = 1, 2(n_i \times n_i)$
r_P = rank of matrix P
s_i = number of streams in category $i = 1, 2$
S = stoichiometric matrix for process $(p \times m)$
$T = Z^THZ(r_z \times r_z)$
v = vector of unmeasured component flow rates corresponding to B_3, combined with unknown extents of reaction $[(n_3 + p) \times 1]$
x = vector of component flow rates corresponding to $B_1(n_1 \times 1)$
Y = matrix, Eq. 8 $[m \times (m - r_P)]$
Z = matrix, Eq. 13 $[r_Y \times (r_Y - r_D)]$
z^a = vector of unit normal variates for test of adjustment a

Greek Letters

δ = vector of adjustments to concentrations $d(n_2 \times 1)$
λ = vector of Lagrange multipliers $(r_z \times 1)$
ξ = vector of extents of reaction $(p \times 1)$
σ_m^2 = variance of M
Σ = variance-covariance matrix:
Σ_1 = component flow rates $x(n_1 \times n_1)$
Σ_c = concentrations $c(n_1 \times n_1)$
Σ_d = concentrations $d(n_2 \times n_2)$
Σ_2 = matrix, Eq. 19 $(n_2 \times n_2)$

APPENDIX: STATISTICAL ANALYSIS

The estimates $\hat{x}$, $\hat{d}$ can be regarded as maximum likelihood estimates, subject to Eq. 15, of x and d, respectively. We wish to deduce the mean and variance of the residual (with $k = 0$ for simplicity)

$$e \triangleq Z^TY^TB_1x \quad \text{(A1)}$$

and of the adjustments a and δ, based on the assumption that x,d are independently normally distributed:

$$\delta x \sim N(0,\Sigma_1) \quad \text{(A2)}$$

$$\delta d \sim N(0,\Sigma_d) \quad \text{(A3)}$$

with

$$\delta x \triangleq x - \hat{x}$$

$$\delta d \triangleq d - \hat{d} \quad \text{(A4)}$$

Let us define

$$T \triangleq Z^THZ \quad \text{(A5)}$$

We note from the definitions in Eq. 12 of D and Eq. 13 of Z, that

$$\frac{\partial Z^T}{\partial d_i} D = -Z^TY^T[0 \ . \ . \ . \ 0B_{2,i}, 0 \ . \ . \ . \ 0] \quad \text{(A6)}$$

where $B_{2,i}$ is the column of B_2 corresponding to d_i and i corresponds to any category 2 concentration. Furthermore,

$$\hat{Z} \triangleq Z(\hat{d}) \quad \text{(A7)}$$

and by definition

$$\hat{Z}^TD(\hat{d}) = 0 \quad \text{(A8)}$$

Then, from Eqs. 9 and 14

$$Z^TY^TB_1\hat{x} = 0 \quad \text{(A9)}$$

Expanding the residual e in a Taylor series to first order, and taking account of Eq. A8

$$e = \hat{Z}^TY^TB_1\delta x + \sum_i \frac{\partial \hat{Z}^T}{\partial d_i} Y^TB_1\hat{x}\, \delta d_i \quad \text{(A10)}$$

Now, with Eqs. A6, 9, and 14

$$e = \hat{Z}^TY^TB_1\delta x + \hat{Z}^TY^TB_2N\, \delta d \quad \text{(A11)}$$

Thus,

$$E(e) = 0 \quad \text{(A12)}$$

$$E(ee^T) = \hat{Z}^TY^T(B_1\Sigma_1B_1^T + B_2N\Sigma_dNB_2^T)Y\hat{Z} \quad \text{(A13)}$$

In the limit, $\Sigma_2 \to N\Sigma_dN$ if the solution is convergent, so that from Eqs. A5 and 31 we obtain

$$E(ee^T) = \hat{Z}^TH\hat{Z} = \hat{T} \quad \text{(A14)}$$

Thus, the residual e is approximately normally distributed as

$$e \sim N(0,\hat{T}) \quad \text{(A15)}$$

The distributions of a and δ can also be approximated from Eqs. 28, 29, and 32, that is

$$a = -\Sigma_1B_1^TYZT^{-1}e \quad \text{(A16)}$$

$$\delta = -N^{-1}\Sigma_2B_2^TYZT^{-1}e \quad \text{(A17)}$$

A Taylor series expansion with $\hat{e} = 0$ gives

$$a = -\Sigma_1B_1^TY\hat{Z}\hat{T}^{-1}e \quad \text{(A18)}$$

$$\delta = -N^{-1}\Sigma_2B_2^TY\hat{Z}\hat{T}^{-1}e \quad \text{(A19)}$$

Thus,

$$E(a) = 0, E(\delta) = 0$$

$$E(aa^T) = \Sigma_1B_1^TY\hat{Z}\hat{T}^{-1}\hat{Z}^TY^TB_1\Sigma_1 \triangleq Q_1 \quad \text{(A20)}$$

$$E(\delta\delta^T) = N^{-1}\Sigma_2B_2^TY\hat{Z}\hat{T}^{-1}\hat{Z}^TY^TB_2\Sigma_2N^{-1} \quad \text{(A21)}$$

If $\Sigma_2 = N\Sigma_dN$

$$E(\delta\delta^T) = \Sigma_dNB_2^TY\hat{Z}\hat{T}^{-1}\hat{Z}^TY^TB_2N\Sigma_d \triangleq Q_2 \quad \text{(A22)}$$

Note that Q_1, Q_2 will be singular if a, δ respectively have more elements than the order of T. Thus, a, δ have degenerate normal distributions.

LITERATURE CITED

Beckman, J. R., "Data Adjustment for Nonreactive Batch or Steady State Processes," *Chem. Eng. Commun.*, 15, 357 (1982).

Britt, H. I., and R. H. Luecke, "The Estimation of Parameters in Nonlinear Implicit Models," *Technometrics*, **15**, 233 (1973).

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 (1983).

Cutting, G. W., "Estimation of Interlocking Mass Balances on Complex Mineral Beneficiation Plants," *Int. J. Mineral Proc.*, **3**, 207 (1976).

Ham, P. G., G. W. Cleaves, and J. K. Lawlor, "Operation Data Reconciliation: An Aid to Improved Plant Performance," *Proc. 10th World Petroleum Congr.*, Bucarest, **4**, 281 (1979).

Hlavacek, V., "Analysis of a Complex Plant—Steady State and Transient Behavior," *Comput. Chem. Eng.*, **1**, 75 (1977).

Hodouin, D., and M. D. Everell, "A Hierarchical Procedure for Adjustment and Material Balancing of Mineral Process Data," *Int. J. Mineral Proc.*, **7**, 91 (1980).

Hodouin, D., T. Gelpe, and M. D. Everell, "Sensitivity Analysis of Material Balance Calculations—An Application to a Cement Clinker Grinding Process," *Powder Technol.*, **32**, 139 (1982).

Iordache, C., R. S. H. Mah, and A. C. Tamhane, "Detection of Gross Errors in Process Data Using Measurement Test," *Paper 43c*, AIChE Spring Meet. Anaheim, CA (1984).

Knepper, J. C., and J. W. Gorman, "Statistical Analysis of Constrained Data Sets," *AIChE J.*, **26**, 260 (1980).

Laguitton, D., "Material Balance of Mineral Processing Flowsheets. FORTRAN Program MATBAL2 Users' Manual," Division Report MRP/MSL 80-33 (IR), CANMET, Ottawa (1980).

Mah, R. S. H., "Design and Analysis of Process Performance Monitoring Systems," *Chem. Proc. Control 2*, Proc. Eng. Found. Conf., Sea Island, GA, *AIChE*, 525 (1981).

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

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

Mular, A. L., et al., "Mass Balance of a Grinding Circuit," *CIM Bull.*, **69**, 124 (Dec., 1976).

Ragot, J., and M. Aubrun, "A Useful Technique for Metallurgical Mass Balances—Applications in Flotation," *Proc. 3rd IFAC Symp. Automation in Mining, Mineral and Metal Processing*, Montreal, 229 (1980).

Romagnoli, J. A., "On Data Reconciliation: Constraints Processing and Treatment of Bias," *Chem. Eng. Sci.*, **38**, 1,107 (1983).

Romagnoli, J. A., and G. Stephanopoulos, "On the Rectification of Measurement Errors for Complex Chemical Plants," *Chem. Eng. Sci.*, **35**, 1,067 (1981).

Smith, H. W., and N. Ichiyen, "Computer Adjustment of Metallurgical Balances," *Can. Inst. Mining Metall. (C.I.M.) Bull.*, **66**, 97 (Sept., 1973).

Smith, H. W., and C. L. Lewis, "Computer Control Experiments at Lake Dufault," *Can. Inst. Mining Metall. (C.I.M.) Bull.*, **62**, 109 (Feb., 1969).

Stanley, G. M., and R. S. H. Mah, "Observability and Redundancy in Process Data Estimation," *Chem. Eng. Sci.*, 36, 259 (1981a).

———, "Observability and Redundancy Classification in Process Networks. Theorems and Algorithms," *Chem. Eng. Sci.*, **36**, 1,941 (1981b).

Vaclavek, V., "Studies on System Engineering III. Optimal Choice of the Balance Measurements in Complicated Chemical Engineering Systems," *Chem. Eng. Sci.*, **24**, 947 (1969).

Vaclavek, V., M. Kubicek, and M. Louchka, "Calculation of Material Balances for Chemical Engineering Systems with Due Allowance for Measurement Errors," *Theor. Found. Chem. Eng.*, **9**, 242 (1976a). (Translation of Teor. Osnovy Khim. Tekhnol. 9, 270 [1975].)

———, "Calculation of Material Balances for Chemical Engineering Systems with Due Allowance for Errors in Measurement Classification of Stream Parameters," *Theor. Found. Chem. Eng.*, **10**, 256 (1976b). (Translation of Teor. Osnovy Khim. Tekhnol. **10**, 281 [1976].)

White, J. W., R. L. Winslow, and G. J. Rossiter, "A Useful Technique for Metallurgical Mass Balances—Applications in Grinding," *Int. J. Mineral Proc.*, **4**, 39 (1977).

Wiegel, R. L., "Advances in Mineral Processing Material Balances," *Can. Metall. Quart.*, 1, 413 (1972).

Manuscript received Aug. 16, 1984, and revision received May 14, 1985.

ERRATA

In the paper titled "Reconciliation of Process Flow Rates by Matrix Projection" by C. M. Crowe, Y. A. Garcia Campos, and A. Hrymak (**29**, p. 881, December, 1983), the following corrections are made:

p. 881, right column, line 10 in Scope: "malfunction of miscalibration" should read "malfunction or miscalibration"

p. 882, right column, line 6 above Eq. 1, Eq. 2, and lines 1, 2 following Eq. 2 as well as Eq. 8 on p. 883: "ϵ" should read "ξ"

p. 884, left column, line 3 under "Detection of Errors": "unknown variance" should read "known variance"

p. 885, add at end of Eq. 35: $= a_j$

p. 887, Table 3, rightmost column heading: "X" should read "χ" twice; and Table 3, line 5, leftmost column: "Straints" should read "Constraints"

p. 888, reference to Kuehn and Davidson: "1967" should read "1961"; and reference to Swenker: "Messergenbnissen" should read "Messergebnissen"