



Data validation for very large scale systems in mineral industry

José Ragot, Didier Maquin, Mohamed Darouach

► To cite this version:

José Ragot, Didier Maquin, Mohamed Darouach. Data validation for very large scale systems in mineral industry. 6th IFAC Symposium on Automation in mining, mineral & metal processing, AMMP'89, Sep 1989, Buenos Aires, Argentina. hal-00293578

HAL Id: hal-00293578

<https://hal.science/hal-00293578>

Submitted on 15 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DATA VALIDATION FOR VERY LARGE SCALE SYSTEMS IN MINERAL INDUSTRY

J. Ragot, D. Maquin, M. Darouach

Laboratoire d'Automatique et de Recherche Appliquée
C.R.V.M. et C.R.A.N.
Rue du Doyen Marcel Roubault - BP 40
54501 Vandœuvre Cédex - FRANCE

Abstract. This communication suggests a method for the hierarchical treatment of measurement balances of static systems through homogeneous multilinear equations. Due to errors caused by sensors, the measurements taken from a physical system are generally noisy and do not verify these balance equations (total or partial flows).

The problem consists in obtaining an estimation of the exact values, which verify the balance equations. The solution is classically obtained by minimizing a quadratic criterion of the error terms, taking into account the balance constraints. A large scale system of nonlinear equations must be solved. Several forms of hierarchical treatment may be used and also suboptimal solutions have been proposed.

The proposed algorithm is based on the mathematical decomposition of the whole system into serial sub-systems. This decomposition is systematic in all cases and allows a direct iteration calculus. The main numerical difficulty is to take into account the coordination factors between each level; after some algebraic manipulations it is possible to eliminate all the Lagrange multipliers of the problem and to express on a very convenient form the coordination factors. Each level of calculus is reduced to a simple projection of the measurements on the constraints space and moreover the same parametrized routine can be used for each level.

An example is given with measurements on a mineralurgical process with quadrilinear models using volumic flowrate, density, granulometric distribution, and chemical analysis. The method has also been tested for processes described by more than one thousand variables.

Keywords. Balance equilibration, error compensation, state estimation, data handling, multivariable system, nonlinear equations, hierarchical calculus.

INTRODUCTION

Data reliability is of fundamental importance for process diagnosis, identification and control. Measurements having large, random or biased errors which go undetected lead to poor control of processes. Detection of such errors is therefore very important, but can only be carried out on the basis of a certain knowledge of the process, of the structure, of the location of the data, and a certain degree of redundancy.

The calculation of the steady-state mass balances in the streams of a mineralurgical process is a basic tool (Hodouin 1986, Laguiton 1980). To calculate these balances, one collects as many measures as practicable of flowrates (volumic or massic), density, concentrations, size distributions in the process streams. These raw measured data are in fact subject to random and possibly gross errors so that they may be inconsistent with the conservation of mass. The data must then be adjusted so that the reconciled values obey the conservation law and some criterion of the weighted measures is minimized.

This reconciliation problem is very classic and illustrate a particular application of the estimation problem. In this work we are interested with steady-state processes described by multilinear mass balance equations; the measurement errors are taken as realization of random variables with normal and zero mean distributions (the case of non zero-mean distribution can be used when the sensors have biased values; in these conditions the bias must be estimated as all other parameters).

The estimation of the true data use the likelihood function maximization in the case where the variance-covariance matrix of the measurement errors is known. This was accomplished originally by Madansky (1959) and Linnik (1963) on a general point of view and by Rosen (1962) and Vaclavek (1969) on a practical point of view.

The approach here is to define a suitable algorithm for the large scale processes when they are described by d-linear constraints.

PROBLEM FORMULATION

The class of d-linear systems is commonly used to describe knowledge models of physical processes under steady state conditions in which it is required to take into account total and partial material flows. The models are said to be structurally exact which is a necessary condition to allow the detection of errors; the general form of model is then:

$$f_j = M \prod_{i=1}^j X_i \quad j = 1, \dots, d \quad (1)$$

- where d is the maximal order of the model,
M is the incidence matrix of the process flowsheet,
X_i is the vector of the true data corresponding to the variable number i,
II is the operator which associate to two vectors a third one such as its components are given by the product of the components of the two vectors (we also note * this product).

By example:

$$f_1 = M \overset{*}{Q} = 0 \quad (2)$$

$$f_2 = M \overset{*}{Q} \overset{*}{p} = 0 \quad (3)$$

can be used to represent the conservation equations for the total volumic flowrate (Q_i) and the conservation in terms of total

massic flowrate ($Q_i \rho_i$). Common applications use quadrilinear forms when we want to take into account volumic flowrate, densities, concentrations and size distribution. However, we consider the general case of d-linear representation.

Let us assume that each flowrate measurement is independent and normally distributed with unknown mean and known variance ; the data validation problem is then reduced to a minimum search problem. The vector \hat{X} of the estimated data \hat{X} is obtained from the measurements X , by minimizing the criterion J :

$$J = \frac{1}{2} \sum_{j=1}^d (\hat{X}_j - X_j)^T V_j^{-1} (\hat{X}_j - X_j) \quad (4)$$

where V_j is the variance-covariance of the measurements errors.

Using Lagrange multipliers, we can solve problem (4) by finding a stationary point (X, λ) of the Lagrangian :

$$L = \frac{1}{2} \sum_{j=1}^d [(\hat{X}_j - X_j)^T V_j^{-1} (\hat{X}_j - X_j) + \lambda_j^T M \prod_{i=1}^j \hat{X}_i] \quad (5)$$

This gives the two equations :

$$V_j^{-1} (\hat{X}_j - X_j) + \sum_{k=1}^d (M \prod_{i=1, i \neq j}^k \hat{X}_i)^T \lambda_k = 0 \quad (6)$$

$$M \prod_{i=1}^j \hat{X}_i = 0 \quad (7)$$

It is clear that the dimension of this system can be very important and moreover the different equations are non-linear ; however, its formulation point out the advantages of a resolution using the principles of decomposition and coordination of the calculus. Other principles of resolution can be applied, in particular the complete linearization of the equations of constraint ; in fact, in the case of d-linear equations it seems that the principles of hierarchical calculus are more powerful and take more advantage of the structure of the equations.

HIERARCHICAL RESOLUTION

In order to point out the principle of this solution we formulate the problem about quadrilinear models which is enough to see the generalization to d-linear systems .

The Lagrangian is defined by :

$$L = \frac{1}{2} \sum_{j=1}^4 [(\hat{X}_j - X_j)^T V_j^{-1} (\hat{X}_j - X_j) + \lambda_j^T M \prod_{i=1}^j \hat{X}_i] \quad (8)$$

For simplicity, we define :

$$M * \hat{X}_i * \hat{X}_j = M_{ij} \quad (9)$$

Finding the stationary point of L gives :

$$V_1^{-1} (\hat{X}_1 - X_1) + M^T \lambda_1 + M_2^T \lambda_2 + M_{23}^T \lambda_3 + M_{234}^T \lambda_4 = 0 \quad (10)$$

$$M \hat{X}_1 = 0 \quad (11)$$

$$V_2^{-1} (\hat{X}_2 - X_2) + M_1^T \lambda_2 + M_{13}^T \lambda_3 + M_{134}^T \lambda_4 = 0 \quad (12)$$

$$M_1 \hat{X}_2 = 0 \quad (13)$$

$$V_3^{-1} (\hat{X}_3 - X_3) + M_{12}^T \lambda_3 + M_{124}^T \lambda_4 = 0 \quad (14)$$

$$M_{12} \hat{X}_3 = 0 \quad (15)$$

$$V_4^{-1} (\hat{X}_4 - X_4) + M_{123}^T \lambda_4 = 0 \quad (16)$$

$$M_{123} \hat{X}_4 = 0 \quad (17)$$

The solution of this strongly nonlinear system can be achieved by decomposition-coordination algorithm. For the i th level, we obtain the estimation of X_i .

We premultiply the equation (10) by MV_1 ; the system can be solved for λ_1 and then gives the estimation \hat{X}_1 as :

$$\hat{X}_1 = [I - V_1 M^T (M V_1 M^T)^{-1} M] [X_1 - V_1 (M_2^T \lambda_2 + M_{23}^T \lambda_3 + M_{234}^T \lambda_4)] \quad (18)$$

We simplify this form with the help of the diagonal matrix Λ_i which the diagonal elements are the components of same order of \hat{X}_i . The expression (9) can be written :

$$M_{ij} = M \Lambda_i \dots \Lambda_j \quad (19)$$

and the estimator \hat{X}_1 :

$$\hat{X}_1 = [I - V_1 M^T (M V_1 M^T)^{-1} M] [X_1 - V_1 \Lambda_2 (M^T \lambda_2 + \Lambda_3 M^T \lambda_3 + \Lambda_3 \Lambda_4 M^T \lambda_4)] \quad (20)$$

Similarly, with notation (19), the equation (12) is written :

$$\hat{X}_2 - X_2 + V_2 \Lambda_1 (M^T \lambda_2 + \Lambda_3 M^T \lambda_3 + \Lambda_3 \Lambda_4 M^T \lambda_4) = 0 \quad (21)$$

Putting (21) into (20) gives :

$$\hat{X}_1 = [I - V_1 M^T (M V_1 M^T)^{-1} M] [X_1 - V_1 \Lambda_2 \Lambda_1^{-1} V_2^{-1} (\hat{X}_2 - X_2)] \quad (22)$$

Similarly, we can obtain :

$$\hat{X}_2 = [I - V_2 M_1^T (M_1 V_2 M_1^T)^{-1} M_1] [X_2 - V_2 \Lambda_3 \Lambda_2^{-1} V_3^{-1} (\hat{X}_3 - X_3)] \quad (23)$$

$$\hat{X}_3 = [I - V_3 M_{12}^T (M_{12} V_3 M_{12}^T)^{-1} M_{12}] [X_3 - V_3 \Lambda_4 \Lambda_3^{-1} V_4^{-1} (\hat{X}_4 - X_4)] \quad (24)$$

$$\hat{X}_4 = [I - V_4 M_{123}^T (M_{123} V_4 M_{123}^T)^{-1} M_{123}] [X_4] \quad (25)$$

In the general case, for the level i , we have :

$$\hat{X}_{i+1} = [I - V_{i+1} M_{1,i}^T (M_{1,i} V_{i+1} M_{1,i}^T)^{-1} M_{1,i}] [X_{i+1} - V_{i+1} \Lambda_{i+2} \Lambda_{i+1}^{-1} V_{i+2}^{-1} (\hat{X}_{i+2} - X_{i+2})] \quad (26)$$

Then each level of calculus use a projection matrix (Crowe 1983), progressively modified by the estimations of the preceding levels. The equations (22) to (25) are directly solved with the measurements as initial choice of the estimations ; the convergence of the calculus is always obtained by a proper choice of a relaxation filter between the different levels. The iterations are stopped when, between two iterations, the numerical values of the estimated do not significantly vary.

SENSITIVITY AND VARIANCE

As the balances equations are nonlinear, it is quite impossible to express an exact formulation of the sensitivity and of the precision of the estimators. In the following section we suggest to use a suboptimal solution obtained by a linearization of the balance equations.

Residues of Balance Equations

The validity of the estimation method depends on the hypothesis about the distribution of the error measurements ; if they are not respected, the validated data have no longer signification. The main fact which can oblige to reject these hypothesis is the presence of faulty data. The detection of faulty data can be done a priori by examination of the residues of balance equations.

The residues of the balance equations, the imbalance vector R_j , are defined as :

$$R_j = M \prod_{i=1}^j X_i \quad j=1, \dots, d \quad (27)$$

We note that measurements are given by :

$$X_i = \hat{X}_i + \varepsilon_i \quad (28)$$

then :

$$R_j = M \prod_{i=1}^j (\hat{X}_i + \varepsilon_i) \quad (29)$$

To express the statistical properties of the imbalance vector, we use the linearized form :

$$R_j = \sum_{k=1}^j M \left[\left(\prod_{i=1, i \neq k}^j \hat{X}_i \right) * \varepsilon_k \right] \quad (30)$$

The mathematical expectation of the measurement errors is taken as zero so that we obtain :

$$E(R_j) = 0 \quad (31)$$

and

$$V(R_j) = \sum_{k=1}^j M \left(\prod_{i=1, i \neq k}^j \hat{X}_i \right) V_k \left(\prod_{i=1, i \neq k}^j \hat{X}_i \right) M^T \quad (32)$$

This expression can be reduced in the case where the variance of the measurements is proportional to their square values :

$$V_k = a \Lambda_k \Lambda_k \quad (33)$$

Then, if in expression (32) we replace the true values (unknown) by their measurements, we obtain as expression of the variance of the imbalance terms :

$$V(R_j) = a j M \left(\prod_{i=1}^j \Lambda_i \right)^2 M^T \quad (34)$$

Finally, we define the residues under a normalized form which allows their relative comparison :

$$R_i^N = \frac{M \prod_{i=1}^j X_i}{\sqrt{a j \left(\prod_{i=1}^j \Lambda_i \right)^2}} \quad (35)$$

Each component of this vector is distributed under a normal distribution (zero mean and unity variance). With the knowledge of the distribution of the residues it is possible to determine, for each of them, the probability to obtain a value less than the observed value. This probability, compared to a critical probability, allows to accept or to reject the hypothesis of fault on the balance residue of node i . The streams connected to this node are suspicious.

Approached Estimator of the True Data

In order to suppress the nonlinear relation between the balance equations we suggest the following variable transformation :

$$\hat{Y}_j = \prod_{i=1}^j \hat{X}_i \quad (36)$$

which put the balance equations under the form :

$$M \hat{Y}_j = 0 \quad j=1, \dots, d \quad (37)$$

As the \hat{Y}_j variables are not measured we define pseudo-measurements Y_j from the measurements X_i :

$$Y_j = \prod_{i=1}^j X_i \quad (38)$$

and pseudo-variances W_j of the pseudo-measurements Y_j from the variances V_j of the measurements X_j :

$$W_j = W_{j-1} \Lambda_j \Lambda_j + V_j \prod_{i=1}^{j-1} \Lambda_i \Lambda_i \quad (39)$$

The Lagrangian (5) is then replaced by :

$$L = \frac{1}{2} \sum_{j=1}^d [(\hat{Y}_j - Y_j)^T W_j^{-1} (\hat{Y}_j - Y_j) + \lambda_j^T M \hat{Y}_j] \quad (40)$$

which appears to be an additive-separable form with respect to the j subscript.

The solution can be expressed by :

$$\hat{Y}_j = [I - W_j M^T (M W_j M^T)^{-1} M] Y_j \quad (41)$$

The equations (39) et (41) give the complete solution of the data reconciliation problem when using the decoupling hypothesis. The reader can consult Fayolle (1987) which gives other techniques of linearization.

Sensitivity and Variance

Taking a variation ΔY_j for the pseudo-measurements, the corresponding variation of the estimations is given by :

$$\Delta \hat{Y}_j = [I - W_j M^T (M W_j M^T)^{-1} M] \Delta Y_j, \quad j=1, \dots, d \quad (42)$$

This equation gives the sensitivity of the variation of a measurement on the estimation of all other variables. It can be used to locate the more sensitive measurements to the reconciliation.

The equation (42) allows also to express the variance ($V()$) of the estimations as :

$$\hat{W}_j = V(\Delta \hat{Y}_j) = E(\Delta \hat{Y}_j \Delta \hat{Y}_j^T) = [I - W_j M^T (M W_j M^T)^{-1} M] W_j, \quad j=1, \dots, d \quad (43)$$

To express the variances of \hat{X}_i we use equation (39) written for the estimated data :

$$\hat{W}_j = \hat{W}_{j-1} \hat{\Lambda}_j \hat{\Lambda}_j + \hat{V}_j \prod_{i=1}^{j-1} \hat{\Lambda}_i \hat{\Lambda}_i \quad j=1, \dots, d \quad (44)$$

from which we can deduce the variance of the estimations \hat{V}_j .

NUMERICAL EXAMPLES

Let us consider (Figure 1) a simple example in the mineralurgical field which can illustrate a group of flotation cells with five nodes (flotation cells) and ten streams.

For each stream we consider the measurements of volumic flowrate Q_i , of density ρ_i , of the proportion t_i of particles belonging interval of size i and of the chemical concentration c_i for mineral specie in this granulometric class.

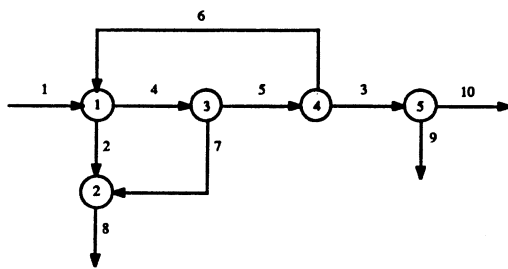


Figure 1. A simple network

For each stream we consider the measurements of volumic flowrate Q_i , of density ρ_i , of the proportion t_i of particles belonging interval of size i and of the chemical concentration c_i for mineral specie in this granulometric class.

The table 1 gives the measurements of these variables. The precisions are expressed by variances of the measurements, which are taken proportional to the square of their magnitude (with 0.1 as proportional factor).

Stream	Volumic flowrate	Massic flowrate	Size distribution	Concentration
1	21.0	1.52	0.61	0.31
2	5.1	1.41	0.59	0.06
3	5.1	1.64	0.82	0.59
4	20.1	1.49	0.59	0.45
5	10.0	1.40	0.72	0.59
6	5.1	1.17	0.50	0.66
7	10.0	1.60	0.50	0.25
8	15.0	1.56	0.51	0.18
9	1.0	1.50	0.93	0.37
10	4.0	1.65	0.84	0.62

TABLE 1 Raw Data

We can appreciate the magnitude of errors and the global incoherence of raw data, by calculating the unbalance terms for each node of the process. In order to compare their relative importance they have been reduced through their respective standard deviation.

	Residue	Normalized residue
r ₁₁	0.90	0.30
r ₁₂	0.10	0.05
r ₁₃	0.10	0.04
r ₁₄	-0.20	-0.16
r ₁₅	0.10	0.15
r ₂₁	0.40	0.06
r ₂₂	-0.20	-0.04
r ₂₃	0.10	0.01
r ₂₄	-0.40	-0.16
r ₂₅	0.30	0.19
r ₃₁	0.10	0.02
r ₃₂	0.30	0.11
r ₃₃	-0.10	-0.02
r ₃₄	-0.20	-0.09
r ₃₅	-0.10	-0.06
r ₄₁	-0.21	-0.10
r ₄₂	0.10	0.16
r ₄₃	0.11	0.05
r ₄₄	-0.10	-0.06
r ₄₅	0.08	0.07

TABLE 2 Residues of Balanced Equations (without bias on measurements)

The residues examination (table 2) don't allow to suspect a balance equation ; the imbalance terms can be considered as normal. To illustrate the necessity of this examination let us consider the data of table 3 ; they have been obtained from table 1 with a bias on stream 3 and variable 3. As it appears in the column of normalized residues, the magnitudes of r_{34} , r_{35} , r_{44} , r_{45} are suspicious and we must try to locate the faulty measurements. In this simple example, a examination of the occurrences of the variables in the different balance equations enables the localization of the bias.

	Residue	Normalized residue
r ₁₁	0.90	0.30
r ₁₂	0.10	0.05
r ₁₃	0.10	0.04
r ₁₄	-0.20	-0.16
r ₁₅	0.10	0.15
r ₂₁	0.40	0.06
r ₂₂	-0.20	-0.04
r ₂₃	0.10	0.01
r ₂₄	-0.40	-0.16
r ₂₅	0.30	0.19
r ₃₁	0.10	0.02
r ₃₂	0.30	0.11
r ₃₃	-0.10	-0.02
r ₃₄	-3.25	-1.27
r ₃₅	3.35	1.63
r ₄₁	-0.10	-0.10
r ₄₂	0.10	0.16
r ₄₃	0.11	0.05
r ₄₄	-2.15	-1.21
r ₄₅	2.13	1.50

TABLE 3 Balance Residues (1 measurement is biased)

The proposed procedure has been applied in order to reconciled the whole data sheet. The table 4 give the obtained estimations.

Stream	Volumic flowrate	Massic flowrate	Size distribution	Concentration
1	20.24	1.55	0.61	0.32
2	5.08	1.42	0.59	0.06
3	5.06	1.61	0.85	0.57
4	20.19	1.49	0.59	0.44
5	10.10	1.40	0.70	0.60
6	5.03	1.18	0.50	0.66
7	10.09	1.59	0.49	0.24
8	15.17	1.53	0.52	0.18
9	1.00	1.50	0.92	0.37
10	4.06	1.64	0.83	0.62

TABLE 4 Reconciled Data

CONCLUSION

A numerical approach has been presented for solving the problem of reconciling flow measurement data in streams of a mineralurgical process. The main idea is to construct a hierarchical algorithm taking into account all the constraints but without keeping into memory the Lagrange parameters. A class of statistic tests has been used to detect gross errors in measurements before the data equilibration ; another statistics are useful to appreciate the sensitivity and the variance of the estimators.

The proposed algorithm, using direct iteration procedure, presents some attractive aspects :

- it gives an optimal solution,
- it is well adapted to large scale problems,
- it only uses simple calculus at each level,
- the coupling between the different constraints do not use auxiliary variables.

In a more general situation, it would be interesting to examine the problem of observability in order to include unmeasured quantities ; actually we have a solution for the bilinear case (Maquin 1989) but we try to extend this result in the case of d-linear system.

REFERENCES

- Hodouin, D., and M.D., Everell (1986). A hierarchical procedure for adjustment and material balancing of mineral process data. Int. J. Min. Proc., **7**, 91-116.
- Laguition, D. (1980). Material balance of mineral processing flowsheets. CANMET Sciences Laboratories. Division Report MRP/MSL 80-33 (IR).
- Fayolle, J. (1987). Validation de données et diagnostic des procédés industriels. Thèse de Doctorat de l'Université de Nancy.
- Maquin, D., M., Darouach, and J., Ragot (1989). Observability and data validation of bilinear systems. Accepted to IFAC Congress AIPAC'89, Nancy, France.
- Crowe, C.M., Y.A., Garcia Campos, and A., Hrymak (1983). Reconciliation of process flowrates by matrix projection. AIChE Journal, **29**, 881-888.
- Madansky, A. (1959). The fitting of straight lines when both variables are subject to errors. J. of the Am. Stat. Assoc., **54**, 552-566.
- Linnik, Y.V. (1963). Méthode des moindres carrés, Dunod.
- Rosen, E.M. (1962). A machine computation method for performing material balances. Chem. Eng. Progr., **58**, N°69.
- Vaclavek, V. (1969). Studies on systems engineering. Chem. Eng. Sci., **24**, 947.