PERIODICA POLYTECHNICA SER. CHEM. ENG. VOL. 38, NOS. 1-2, PP. 7-17 (1994)

MEASUREMENT DATA ANALYSIS
A General Approach on Balance Equation and Maximum Likelihood Estimation Basis

Gedeon A. ALMÁSY

Amerikai út 72.
H-1145 Budapest, Hungary

Received: June 15, 1994

Abstract

A general theoretical background for Measurement Data Analysis, including consistency check, state estimation (called also Balancing or Measurement Error Reconciliation) and Gross Error (GE) Identification is set up on the basis of a set of balance equations or other linear constraints. GE Situations are defined and the use of Maximum Likelihood principle is proposed to estimate both the actual GE situation and the state variables. The essential difference between the use of linear dynamic models and balance equations is shown.

Keywords: measurement error, measurement error reconciliation, gross error identification, balance equations.

1. Introduction

On Measurement Data Analysis (MDA) we mean Consistency Test, State Estimation, called also Measurement Error Reconciliation or Balancing, and Gross Error Identification together. These activities are discussed mostly as independent problems. It is shown here that all these have common *theoretical bases* and can be solved by a single estimation algorithm. The *balance equations* and the *Maximum Likelihood Estimation* are proposed as the common bases of analysis. This approach to MDA can be applied both for stationary and transient states; the first is called static, the second dynamic. The static problem is solved; the problem of dynamic analysis is rather outlined yet.

This paper is a synopsis of author's earlier works and presents a general theoretical background of MDA problems. It is no broad survey on the relevant literature given here; it can be found in author's earlier papers (ALMÁSY, 1990), (ALMÁSY – UHRIN, 1993, 1994) or in others, e. g. that of KAO – TAMHANE – MAH (1992). Only literature directly related to the topic is cited.

2. Balance Equations

Balance equations express the *laws of conservation* of physics, or other *rules of conservation,* in the way that a certain linear combination of the *balance variable vectors* the vector of the so-called *balance-ruled elements* is a constant, or changes only due to streams entering or leaving the system concerned. Balance equations have to be formulated somewhat differently for static and for dynamic systems. Dynamic balance equations necessarily include also the static case.

2.1. Static Balance Equations

Let the number of the balance variable vectors be n. Let the vector of the true values of balance variables be denoted by $\xi \in \Xi$, $\Xi \subset \mathrm{R}^n$, that of those burdened with measurement errors $\mathbf{x} \in \mathrm{R}^n$, and the vector of measurement errors $\mathbf{d} \in \mathrm{R}^n$. Note that in this context element means not necessarily chemical elements but also other physical substances for which any rule of conservation is valid.

The balance equations are linear constraints of the form

$$\mathbf{A} \cdot \xi - \mathbf{b} = \mathbf{0} . \tag{1}$$

with constant matrix parameter $\mathbf{A} \in \mathrm{R}^{p \cdot n}$ of rank p, $p < n$, and vector parameter $\mathbf{b} \in \mathrm{R}^p$, where p is the number of balance equations, so that

$$\Xi = \{\xi \in \mathrm{R}^n : \mathbf{A} \cdot \xi - \mathbf{b} = \mathbf{0}\} .$$

The practical meanings of matrix $\mathbf{A}$ and $\mathbf{b}$ are assumed to be known and are not explained here.

Let us denote the vector of *balance inconsistencies* or *imparities* by $\mathbf{y} \in \mathrm{R}^p$. With the measured values $\mathbf{x}$ *Eq.* (1) becomes

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} - \mathbf{b} . \tag{2}$$

From the above, with

$$\mathbf{d} = \mathbf{x} - \xi ,$$

the vector of balance inconsistencies becomes

$$\mathbf{A} \cdot \mathbf{d} = \mathbf{A} \cdot (\mathbf{x} - \xi) = \mathbf{A} \cdot \mathbf{x} - \mathbf{b} = \mathbf{y} . \tag{3}$$

Because measurement errors are of stochastic nature, vector $\mathbf{d}$ and consequently also $\mathbf{y}$ and $\mathbf{x}$ are random variables. A basic assumption, applied in this paper is that $\mathbf{d}$ is independent of ξ.

Also unmeasured variables can be involved into balances. But, since balances can be reduced easily to the form of *Eq.* (1) even in such cases, if it is possible at all, these are not discussed here explicitly.

2.2 Dynamic Balance Equations

Dynamic balance equations express the laws of conservation, similarly to the static balance equations:

$$\mathbf{A} \cdot \xi^{t-1} - (\eta^{t} - \eta^{t-1}) = \mathbf{0}\,, \tag{4}$$

where $\eta^t \in \mathrm{R}^p$ is the true amount of present elements, called *inventory variables,* at time $t \in T \subseteq \mathrm{R}$. The variable $\xi^t \in \mathrm{R}^n$, $t \in T$ is the true value vector of the so-called *flow variables.*

To the dynamic treatment also some knowledge about the inlet and outlet streams is necessary. It seems to be the simplest way to specify the dynamic behaviour of the flow variables as a self- correlated Gaussian stochastic vector process with incremental variance $\mathbf{V}_r$, i. e. such for which

$$\xi^{t} - \xi^{t-1} = \mathbf{r}^{t-1}\,, \qquad \mathbf{r}^{t} \sim N_n(\mathbf{0}, \mathbf{V}_r)\,. \tag{5}$$

It is more complicated to compute the balance inconsistency vector in the dynamic case.

There is a hereditary error already in the estimate of η^{t-1} itself so that the inconsistency cannot be computed like that in the case of static error, applying to *Eq.* (3), but it is defined by *Eq.* (4) itself. It was shown in an earlier paper, how the Kalman filter can be applied to estimate the balance inconsistency and its variance (ALMÁSY, 1990). Others about dynamic balancing were presented in author's another paper (ALMÁSY, 1986).

3. The Measurement Error Model

3.1. Total Error

For a correct MDA an *exactly defined* measurement error model is necessary. It is postulated that the Total Measurement Error (TE) is a sum of an Ordinary Error (OE) and of a Gross Error (GE) and that OE-s are considered to be *permanently present,* while GE-s as *occasional* additive random components of measurement errors:

$$\mathbf{d}^{t} = \mathbf{b}^{t} + \mathbf{g}^{t}\,. \tag{6}$$

Here $\mathbf{d}^t \in \mathrm{R}^n$ is the random vector variable of the TE, $\mathbf{b}^t \in \mathrm{R}^n$ that of OE-s, and $\mathbf{g}^t \in \mathrm{R}^n$ that of the GE-s. While $\mathbf{b}^t$ is unavoidable, $\mathbf{g}^t$ is only occasional. *Eq.* (6) shall be interpreted so that in the absence of GE $\mathbf{g}^t$ is an exactly zero vector. (For shortness, superscript t will be omitted in the forthcoming.) The occurrence of GE-s is a parameter of the distribution of the total error; the values of GE are random.

Let us assume that the distributions of $\mathbf{b}$ and $\mathbf{g}$ are known. Denote the density function of $\mathbf{b}$ by $\mathbf{f}_b(\bullet)$ and that of $\mathbf{g}$ by $\mathbf{f}_g(\bullet)$. As it is known, the density function of a sum of two random variables is their convolution:

$$\mathbf{f}_d(\bullet) = \mathbf{f}_b * \mathbf{f}_g(\bullet) = \int_{\mathsf{R}^*} \mathbf{f}_b(\bullet - \mathbf{v}) \cdot \mathbf{f}_g(\mathbf{v}) \mathrm{d}\mathbf{v} \,. \tag{7}$$

3.2. Ordinary Error

OE-s are considered usually as n-dimensional, independent, self-uncorrelated $\mathbf{0}$ mean, $\mathbf{V}_b$ variance Gaussian random vector processes. This model was used in the simulation experiments.

3.3. Gross Error

A number of works have been published with the explicit or implicit assumption that GE-s themselves are the unknown additive measurement biases, i. e., shift parameters of the OE distribution. All algorithms, referred earlier, that apply rejecting measurements sequentially and computing the residual error variances from the remaining (RIPPS, 1965, and others), can be regarded as applying this idea. The problem concerning those algorithms is that they introduce the mean of OE as an additional free parameter for GE but its estimation is not done according to the disciplines of mathematical statistics.

3.4. Gross Error Situations

There are generally more than a single possible source of GE. In order to treat this fact correctly, the concept of *GE situations* is introduced. A GE situation is defined as an event when besides an OE there occurs a GE as an additive *random variable of given distribution;* different GE situations with different distributions. Let the unstructured set of all possible GE situations be represented by S with elements s. Thus, the estimation of sources of GE according to this formulation leads to a choice from the elements $s \in S$. Even if s is interpreted as a set of non-negative integer indices its numeric values do not carry any other information but distinguishing the different situations. Distributions of all $\mathbf{g}_s$-s, belonging to different s-s, represent different GE situations; let their density functions be denoted by $\mathbf{f}_{g_s}(\bullet)$, $s \in S$.

Applying the above concept and notation, the TE is considered as the sum of OE and occasionally one of the GE-s, if GE is present:

$$\mathbf{d} = \mathbf{b} + \mathbf{g}_s, \; s \in S\,. \tag{8}$$

The density function of this sum is their convolution:

$$\mathbf{f}_{d_s}(\bullet) = \mathbf{f}_b * \mathbf{f}_{g_s}(\bullet)\,, \tag{9}$$

where subscript s means that the GE corresponds to the situation s.

The no-GE situation shall be included into S by, say, $s = 0$ with the degenerated multidimensional Dirac delta distribution at vector $\mathbf{0}$ (Dirac-delta at zero):

$$\mathbf{f}_{g_0}(\bullet) = \delta(\bullet, \mathbf{0})\,.$$

In the case of the no-GE situation the TE is equal to the OE and

$$\mathbf{f}_{d_0}(\bullet) = \mathbf{f}_b * \delta(\bullet, \mathbf{0}) = \mathbf{f}_b(\bullet)\,. \tag{10}$$

This latter expresses that all samples taken from $\mathbf{g}_0$ are equal to $\mathbf{0}$ so that adding $\mathbf{g}_0$ to $\mathbf{b}$ does not change the distribution of $\mathbf{b}$.

The most general formulation of the estimation problem is to consider OE-s and GE-s as dependent on each other. It seems to be. however, a practical restriction in the problem formulation that each GE is regarded as independent of any OE.

The cardinality of S is considered as finite. If there are no special reasons, it seems to be a reasonable choice to include two GE situations for each variable, representing a positive and a negative GE, besides the no-error situation. This choice corresponds to the practical assumption that there is either no GE or there is only a single one, either positive or negative in the measurements. With this specification of S, the cardinality is $2 \cdot n + 1$. Neglecting simultaneous independent GE-s seems to be appropriate because of their very low probability. Other GE situations can be specified as well if there is any reason for them, mostly, if there is any common source for several GE-s.

Note that the concept of GE situations is akin to that applied by ROSENBERG – MAH – IORDACHE (1987), called the candidate set of streams. We emphasize here the fact that the random occurrence of two or more independent GE-s is extremely rare, assumed that the technology is kept in good repair. Lots of computation time can be saved in this way.

Author's opinion is that the specification of GE situations is the duty of the technical management, on the basis of engineers' knowledge about the process. Some considerations concerning the distribution of GE were listed in an earlier paper (ALMÁSY, 1993).

4. The Maximum Likelihood State Estimation

4.1. The Likelihood Function Values for a Measured State

According to the basic definition given previously, the GE Identification problem is regarded as the simultaneous estimation of the presence and the value of GE as a random variable. The Maximum Likelihood (ML) approach was applied to GE Identification also by TJOA – BIEGLER (1991) but for a less general problem.

The set of parameters will be the true value vector $\xi \in \Xi \subset \mathbb{R}^n$, as the shift parameter of the distribution. This will be explicitly written among the parameters of distributions. All other parameters of the density function of TE will be concentrated within the notation s of the GE situation. According to the above, the set of parameters to be estimated becomes the direct product $\Xi \times S$.

Let us denote the density function of the random variable $\mathbf{x}$ in the s GE situation by $\mathbf{f}_s(\mathbf{x};\xi)$. Thus, the ML estimate $\widehat{s} \in S$ of s and $\widehat{\xi} \in \Xi$ of ξ is that where

$$\mathrm{l}\left(\widehat{\xi},\, \widehat{s};\, \mathrm{x}\right) = \max_{\xi\in\Xi, s\in S} \left(\mathrm{l}\left(\xi,\, s;\, \mathrm{x}\right)\right) = \max_{\xi\in\Xi, s\in S} \left(\mathbf{f}_s\left(\mathrm{x};\, \xi\right)\right), \qquad \mathrm{x} \in \mathbb{R}^n. \quad (11)$$

So, the solution of the estimation problem needs the maximization of *Eq.* (11) simultaneously on s and ξ over $\Xi \times S$. Since to each s corresponds a distribution, including its parameters, $\widehat{s}$ means the optimal choice from all possible GE situations S, with optimally chosen $\widehat{\xi}_s \in \Xi$ for each s.

4.2. Maximum Search

Because S is an unstructured set, no other general way to maximize the likelihood function is known than some maximum search on $\mathbf{x}$ at each s, followed by a total enumeration according to $s \in S$. Thus, the cardinality of S has to be kept low in practice. If both GE and OE are Gaussian, $\widehat{\xi}_s$, maximizing the likelihood function at each fixed s, can be computed relatively easily by a quadratic form computation with a prefabricated coefficient matrix for each s. For the techniques of computation and examples see ALMÁSY (1994).

5. Measurement Data Analysis

5.1. Static Measurement Data Analysis

With the concepts introduced above, all three aims of MDA can be performed simultaneously.

5.1.1. Measurement consistency check and GE Identification

According to the foregoing, consistency of measurements means that the no-GE situation is of the highest probability. If the likelihood function value is higher in any other situation, that situation is regarded as the ML estimate of the GE. Note that if there are more GE situations with nearly equal likelihood function values, all those shall be regarded as possible sources of the observed GE.

5.1.2. State estimation

State estimation can be done, applying the mean value vector and variance matrix corresponding to the estimated GE situation $\widehat{s}$, even if other than the no-GE situation is found as most likely. The ML state estimate is $\widehat{\xi}_{\widehat{s}}$, i. e. $\widehat{\xi}_s$ with $s = \widehat{s}$. Naturally, the estimate is most reliable in the no-GE case.

5.2. About Dynamic Measurement Data Analysis

The principle of MDA on the balance equation basis and ML principle is not constrained to the static case, but it has not been worked out in details and tested by examples for the dynamic case yet. It is possible to express the likelihood function value for each dynamic GE situation and to choose the ML situation and state as the estimate. Similar ideas, but not on the ML estimation principle, have been proposed using the dynamical linear process model by several authors where instead of the likelihood function values the residual variances have been taken into consideration. This technique is mostly recommended in control theory. These algorithms, called Fault Detection and Isolation, were summarized by FRANK (1989). The problem is complicated, compared to the steady state, not only because it needs Kalman filtering instead of solving a constrained minimisation of a quadratic form, but also because extraordinary events are not only the GE-s in measurements but also because extraordinary events are not only the GE-s in measurements but also sudden changes in the inputs, actuator faults, etc. These can be taken into account in the theory, but the number of possible situations increases rapidly if all such events should be considered.

6. Comparing the Analysis Based on the Dynamic Model and that on the Balance Equation Basis

MDA on balance equation basis differs both in its theoretical bases and its technique from the so-called 'sensor validation' or 'fault detection and isolation, based on structural redundancy', the topics of control theory.

6.1. Decomposition

Showing the difference, we constrain ourselves to linear models. Let the linear discrete time state space model of a system be

$$\mathbf{x}^t = \Phi\cdot\mathbf{x}^{t-1}+\mathbf{G}\cdot\mathbf{u}^{t-1}\,, \qquad \mathbf{x}^t, \mathbf{x}^{t-1} \in \mathrm{R}^m\,, \qquad \mathbf{u}^{t-1} \in \mathrm{R}^e\,, \qquad t-1,\, t \in T\,. \tag{12}$$

If the process undergoes some rules of conservation, some elements of a one-to-one linear transform of the variable $\mathbf{x}_t$ must not change in time if the input vector $\mathbf{u}_t$ is identical to $\mathbf{0}$ for all $t \in T$. The condition of this is that state transition matrix Φ has as many unit eigenvalues as many elements undergo independent laws of conservation in the process. This means that there exists such a similarity transformation of Φ that

$$\Phi = \mathbf{T}_l \cdot \begin{pmatrix} \Lambda_d & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \cdot \mathbf{T}_r \tag{13}$$

with

$$\mathbf{T}_l \cdot \mathbf{T}_r = \mathbf{T}_r \cdot \mathbf{T}_l = \mathbf{I}\,.$$

Λ_d is a $d \times d$ block-diagonal matrix with all its elements or blocks in its diagonal corresponding to the non-unit eigenvalues. $\mathbf{T}_l$ and $\mathbf{T}_r$ are non-singular $n \times n$ matrices composed of the left and right eigenvectors of Φ.

Therefrom, with

$$\mathbf{z} = \mathbf{T}_l \cdot \mathbf{x}$$

we obtain

$$\begin{pmatrix} \mathbf{z}_d^t \\ \mathbf{z}_b^t \end{pmatrix} = \begin{pmatrix} \Lambda_d \cdot \mathbf{z}_d^{t-1} \\ \mathbf{I} \cdot \mathbf{z}_b^{t-1} \end{pmatrix}$$

and the original dynamic model decomposes into a *'dynamic'* and *'balance* submodel

$$\mathbf{z}_d^t = \Lambda_d \cdot \mathbf{z}_d^{t-1} \tag{14}$$

and

$$\mathbf{z}_b^t = \mathbf{z}_b^{t-1} \,. \tag{15}$$

Eq. (14) describes the *dynamics* of the system and has no unit coefficient modes. So, it is suitable to design the control of the process where unit eigenvalues would cause troubles.

In turn, *Eq.* (15) expresses the balances and shows that if the discrete time process contains $b = n - d$ elements for which independent rules of conservation are valid, the state transition matrix Φ of its model has just b number of unit eigenvalues. This statement is also inversely true. If the state transition matrix Φ of a process model has certain number of unit eigenvalues, the process contains exactly the same number of independent elements for which laws or rules of conservation are valid. Thus, *Eq.* (15) is good to test and analyse measurements, as described previously.

Nothing changes if the process has a non-zero input in *Eq.* (12). Only the coefficient matrix $\mathbf{G}$ has to be transformed accordingly. The product

$$\mathbf{T}_l \cdot \mathbf{G} = \mathbf{A}$$

gives just the coefficient matrix of the balance equations.

6.2. Practical Consequences

There are essential differences between the two approaches. Static balancing applies, at least in its strict sense, only exact and unquestionable knowledge about the process (apart from that of the measurement error distribution) but does not take into consideration the time correlation of data or any other information on process behaviour. Thus, the verification of measurements is not complete by balancing; only errors influencing the balance can be detected in this way.

Dynamic balancing needs also the knowledge of the incremental variance matrix of the input and output variables, but all other matrix coefficients of the process are exact.

On the other hand, methods of Fault Detection and Isolation suppose the exact knowledge of the process dynamics and work sufficiently quickly only if it is linear. But neglecting of non-linearities in process engineering is dangerous and can strongly mislead estimation. Sometimes the model accuracy is less than that of the measurements so that state estimation on this basis would cause higher error than unfiltered measured data are subject to.

7. Conclusion

It is shown that all MDA problems, including test for consistency, state estimation and GE Identification, are essentially on the same theoretical basis and are proposed to be performed simultaneously, applying the balance equations and the ML principle for estimation.

Simulation experiments have been done with good results for the static problem but they are not presented here for shortage of space. The analogous treatment for the dynamic case is a proposal only; it has to be worked out in detail in the future.

Acknowledgement

This work has been supported by the Hungarian Science Foundation grant OTKA 1991/178.

References

ALMÁSY, G. A.: Principles of Dynamic Balancing. *AICHE Journal,* Vol. 36, pp. 1321–1330 (1989).

ALMÁSY, G. A. – UHRIN, B.: Principles of Gross Error Identification by Maximum Likelihood Estimation. *Hungarian Journal of Industrial Chemistry,* Vol. 21, pp. 309–317 (1993).

ALMÁSY, G. A. – UHRIN, B.: A Gross Measurement Error Identification Algorithm on the Maximum Likelihood Principle. *Hungarian Journal of Industrial Chemistry,* Vol. 22, pp. 153–160 (1994).

KAO, CH. SH. – TAMHANE, A. C. – MAH, R. S. H.: Gross Error Detection in Serially Correlated Process Data (Parts 1 and 2). *Ind. Eng. Chem. Res.,* Vol. 29. pp. 1004–1012 (1990) and Vol. 31, pp. 254–262 (1992).

ALMÁSY, G. A.: On the Possibilities of Dynamic Observation Failure Indication on Balance Equation Basis. *Proc. 25th Conference on Decision and Control,* Athens, Greece, 1986.

RIPPS, D. L.: *Chem. Engng. Progress, Symp.* Ser. 55. pp. 8–16 (1965).

ROSENBERG, J. – MAH, R. S. H. – IORDACHE, C.: Evaluation of Schemes for Detecting and Identifying Gross Errors in Process Data. *Ind. Eng. Chem. Res.*, Vol. 26. pp. 555–573 (1987).

TJOA, I. B. – BIEGLER, L. T.: Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Non-linear Systems. *Computers Chem. Engng.*, Vol. 15, pp. 679–690 (1991).

FRANK, P. M.: Evaluation of Analytical Redundancy for Fault Diagnosis in Dynamic Systems. *Prepr. of IFAC, IMACS, IFORS Symposium AIPAC'89*, Nancy, France, 1989, Vol. 1, pp. 7–19.