Available online at www.sciencedirect.com
ScienceDirect
Chemical Engineering and Processing 46 (2007) 1251–1263
Chemical Engineering and Processing
Process Intensification
www.elsevier.com/locate/cep

Enhancing dynamic data reconciliation performance through time delays identification

Ignacio Yélamos, Carlos Méndez, Luis Puigjaner *

Universitat Politècnica de Catalunya, Chemical Engineering Department, Av. Diagonal 647, 08028 Barcelona, Spain

Received 16 October 2006; accepted 16 October 2006
Available online 30 November 2006

Abstract

In order to improve the performance of data reconciliation methods, an efficient Genetic algorithm (GA) for determining time delays has been developed. Delays are identified by searching the maximum correlation among the process variables. The delay vector (DV) is integrated within a dynamic data reconciliation (DDR) procedure based on Kalman filter through the measurements error model. The proposed approach can be satisfactorily applied not only off-line but also on-line. It was firstly validated in a dynamic process with recycles and feedback control loops. Then, the methodology was successfully applied to a highly non-linear and complex challenging control case study, the Tenessee Eastman benchmark process, demonstrating its robustness in complex industrial problems. This case study required to implement an extended Kalman filter to deal with the existing non-linearities.

Keywords: Genetic algorithm; Time delay identification; Dynamic data reconciliation; Kalman filter

1. Introduction

Data reconciliation (DR) is a model-based filtering technique that attempts to reduce the inconsistency between measured process variables and a process model. Thus, by decreasing this inconsistency, measurement errors are reduced and more accurate process variable values are obtained, which then can be used for process control and optimization. Nevertheless, the performance of DR can be seriously affected by the presence of any event that increases this inconsistency such as modeling errors, gross-errors and/or delays in sampling data. In such situations, an error in the estimation can be generated by forcing a matching of the process model and measured signals that are incompatible. To deal with the presence of gross-error(s) several approaches have been presented in the literature [1], covering steady and dynamic systems as well as different kinds of errors (e.g. bias, process leak). However, less attention has been devoted to manage situations in which delays are present. Since time delays appear as a time variable, they introduce non-linearities even if the process is a linear one.

Correlation measures have been previously used in order to determine process delays. Wachs and Lewin [2] proposed an algorithm that applies relative shifts between process signals within a process data matrix in order to find out the position in which the correlation among variables is maximum. The optimal shifts are those that minimize the determinant of the associated correlation matrix. Such mechanism was applied for improving resolution of a PCA based fault diagnosis system. When the number of process variables n is large or the maximum delay $d_{\max}$ is high in terms of sampling time, exhaustive search of the maximum correlation by direct data matrix manipulation is impracticable due to the unmanageable combinatorial size (it involves $(d_{\max})^n$ determinant calculations). Wachs and Lewin [2] realized this problem and proposed a new algorithm that reduces the problem size to $d_{\max} \times s \times (n - s)$, where s is the number of inputs. However, this algorithm assumes that the output variables are correlated among themselves with no delays present and that the process inputs are independent.

Apart from such limitation, the reduced but exahustive searching strategy maintained in the methodology, makes it unsuitable for actual industrial problem involving continuous

Abbreviations: CPU, central processing unit; DR, data reconciliation; DDR, dynamic data reconciliation; EKF, extended Kalman filter; GA, genetic algorithm; KF, Kalman filter; PRD, process related delay; SRD, sensor related delay; TDI, time delay identification

* Corresponding author. Tel.: +34 93 401 66 78; fax: +34 93 401 09 79.
E-mail address: Luis.Puigjaner@upc.edu (L. Puigjaner).

0255-2701/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cep.2006.10.013

(not discrete) time delays. Since real industrial problems have to deal with continuous operation times, such delays identification methodology becomes infeasible in the on-line implementations because of the CPU time limitations. In that sense, more efficient and systematic approaches should be developed in order to tackle with the real time delays identification. In this work an efficient time delay identification (TDI) procedure based on GA optimization is presented for determining all the existing delays with respect to a reference process signal (e.g. the most or the least delayed process signal). This information is then introduced within the dynamic data reconciliation (DDR) procedure by means of the measurements error model. An academic dynamic case study with recycles and a challenging industrial control problem reported in the literature, the Tennessee Eastmant benchmark [3] case study have been selected to illustrate the advantages of the DDR approach. The main goal of this work is to properly combine DDR and TDI techniques in such a way that data/model delay mismatches can be reduced as much as possible.

2. Data reconciliation

DR is a technique that takes advantage of the redundancy between the process model and the measurements model (Eq. (1)). Therefore, the existence of both models is a prerequisite to reconcile redundant measured data. The measurements model is constructed by assigning to each one of the n measured process variables a normal probability distribution around its true value. In the absence of gross-errors the measurements model can be written as:

$$y_{t_k} = y^*_{t_k} + \varepsilon_{t_k}, \quad y_{t_k} \in \Re^n \tag{1}$$

where y is the $n \times 1$ measurements vector, y^* the $n \times 1$ vector of unknown true values, and ε stands for the $n \times 1$ vector of random measurement errors whose expected value is the null vector and has a known positive definite variance matrix (measurements are assumed to be independent).

Additional information must be introduced through the process model equations (constraints). The process model is in general represented by a set of differential equation constraints:

$$f\left(\frac{d\hat{y}_{t_k}}{dt}, \hat{y}_{t_k}, \hat{\delta}_{t_k}, \hat{\theta}_{t_k}\right) = 0, \quad f \in \Re^f \tag{2}$$

algebraic equality constraints:

$$g(\hat{y}_{t_k}, \hat{\delta}_{t_k}, \hat{\theta}_{t_k}) = 0, \quad g \in \Re^g \tag{3}$$

and inequality constraints:

$$h(\hat{y}_{t_k}, \hat{\delta}_{t_k}, \hat{\theta}_{t_k}) \le 0, \quad h \in \Re^h \tag{4}$$

where $\hat{y}_{t_k}$ represents the values of the estimated vector at discrete time t_k and $\hat{\delta}_{t_k}$ and $\hat{\theta}_{t_k}$ are the $p \times 1$ vector of estimated unmeasured variables and the $q \times 1$ vector of adjusted model parameters at discrete time t_k, respectively.

The adjustment of measurements in order to compensate random errors usually leads to a constrained optimization problem. In most applications, the objective function (OF) to be minimized is a weighted least-squares of the differences from the measured values subject to the process model constraints.

$$\mathrm{OF}(y, \hat{y}, \sigma) = \sum_{t_k=0}^{K} [\hat{y}_{t_k} - y_{t_k}]^T \sum\nolimits^{-1} [\hat{y}_{t_k} - y_{t_k}] \tag{5}$$

where $\sum$ is a diagonal matrix $\sum = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ and σ_i^2 is the variance of the ith measured variable.

Different approaches have been proposed to solve this optimization problem [4,5]. If the process does not show high non-linearities and no inequality constraints are present, Kalman filter (KF) technique can be efficiently used [6].

2.1. Kalman filter

The Kalman filter (KF) is an efficient recursive filter which estimates the state of a dynamic system from a series of incomplete and noisy measurements, assuming that the true current state evolves from the previous one. The algorithm uses the following state-space and measurements models, respectively:

$$x_{t_k} = A_{t_k} x_{t_{k-1}} + B_{t_k} u_{t_k} + w_{t_k} \tag{6}$$

$$y_{t_k} = H_{t_k} x_{t_k} + v_{t_k} \tag{7}$$

where t_k represents a sample time, x_{t_k} is the n_x dimensional vector of state variables, u_{t_k} the n_u dimensional vector of manipulated input variables and y_{t_k} is the n_y dimensional vector of measured variables. The state transition matrix A_{t_k}, the control gain matrix B_{t_k} and the observation matrix H_{t_k} are matrices of an appropriate dimension and if their coefficients are time independent the subscript t_k can be dropped.

The Kalman filter assumes errors in the process model and in the measured data. The process noise w_{t_k} represents errors in the state transition model. This noise is assumed to be white, with zero mean and a variance Q. v_{t_k} represents a measurement noise with a variance R. Using process measurements, the error covariance matrix P_{t_k/t_k} associated with the estimated state vector $\hat{x}_{t_k/t_k}$ is updated as follows:

$$P_{t_k/t_k} = (I - K_{t_k} H_{t_k}) P_{t_k/t_{k-1}} \tag{8}$$

where K_{t_k} is the Kalman filter gain given by:

$$K_{t_k} = P_{t_k/t_{k-1}} H^T_{t_k} (H_{t_k} P_{t_k/t_{k-1}} H^T_{t_k} + R_{t_k})^{-1} \tag{9}$$

Then, the process expected measurements and the model current state estimation can be evaluated by means of Eqs. (10) and (11), respectively.

$$\hat{y}_{t_k} = y_{t_k} - H_{t_k} \hat{x}_{t_k/t_{k-1}} \tag{10}$$

$$\hat{x}_{t_k/t_k} = \hat{x}_{t_k/t_{k-1}} + K_{t_k} \hat{y}_{t_k} \tag{11}$$

where $\hat{x}_{t_k/t_k}$ is the model state prediction and $\hat{y}_{t_k}$ is the measurements model prediction. $y_{t_k} - H_{t_k}\hat{x}_{t_k/t_{k-1}}$ is called the residual or innovation and reflects the discrepancy between the predicted measurements $H_{t_k}\hat{x}_{t_k/t_{k-1}}$ and the real measurement y_{t_k}. The gain matrix K_{t_k} is a factor that allows to weight more or less