Acta Automatica Sinica
Vol. 35, No. 6 June, 2009

Available online at www.sciencedirect.com
ScienceDirect

# A Support Vector Regression Approach for Recursive Simultaneous Data Reconciliation and Gross Error Detection in Nonlinear Dynamical Systems

MIAO Yu[1] SU Hong-Ye[1] CHU Jian[1]

**Abstract** The quality of process data in a chemical plant significantly affects the performance and benefits gained from activities like performance monitoring, online optimization, and control. Since many chemical processes often show nonlinear dynamics, techniques like extended Kalman filter (EKF) and nonlinear dynamic data reconciliation (NDDR) have been developed to improve the data quality. Recently, the recursive nonlinear dynamic data reconciliation (RNDDR) technique has been proposed, which combines the merits of EKF and NDDR techniques. However, the RNDDR technique cannot handle measurements with gross errors. In this paper, a support vector (SV) regression approach for recursive simultaneous data reconciliation and gross error detection in nonlinear dynamical systems is proposed. SV regression is a compromise between the empirical risk and the model complexity, and for data reconciliation it is robust to random and gross errors. By minimizing the regularized risk instead of the maximum likelihood in the RNDDR, our approach could achieve not only recursive nonlinear dynamic data reconciliation but also gross error detection simultaneously. The nonlinear dynamic system simulation results in this paper show that the proposed approach is robust, efficient, stable, and accurate for simultaneous data reconciliation and gross error detection in nonlinear dynamic systems within a recursive real-time estimation framework. It can also give better performance of control.

**Key words** Support vector regression, data reconciliation, gross error detection, nonlinear estimation

Process data measurements are important for model fitting, process monitoring, control, optimization, and management decision making. Unfortunately, process data measurements usually contain two types of errors, random and gross, which severely impact the effects of process monitoring, control, optimization, and management decision making. Meanwhile, chemical processes usually exhibit nonlinear dynamics. So, online estimation of the process states and removal of random and gross errors from measurements are critical for the application of process monitoring, control, and optimization on dynamic nonlinear processes. This task is termed dynamic data reconciliation (or dynamic data rectification)[1−2], which is especially challenging when the state and/or measurement functions are highly nonlinear.

In order to ameliorate the effect of random errors, several different estimation methods have been proposed. For linear dynamic systems, the Kalman filters (KF) give optimal estimates in the presence of measurement uncertainties. Further, extended Kalman filters (EKF) were developed for nonlinear systems, which were based on linearizing the nonlinear equations and applying the Kalman filter to update equations to the linearized system. The advantages of the KF and EKF and their variants lie in their predictive-corrective form and the recursive nature of estimation, which allows for rapid estimation in real-time. Since the KF and EKF are not specifically designed to detect and remove outliers, a probabilistic formulation was proposed[2], which combined the EKF with the expectation-maximization (EM) algorithm to attain the rectified measurements. However, the KF and all its variants cannot take into account bounds on process variables or algebraic constraints, leading to failure of the EKF in many processes[3]. Furthermore, when the states and/or measurement equations are highly nonlinear, KF and all its variants give unsatisfactory state estimates.

An alternative approach called the recursive nonlinear dynamic data reconciliation (RNDDR) has been proposed[4], which is an extension of EKF. So, the RNDDR is preferable for online applications. Moreover, RNDDR could take account of bounds and algebraic constrains for state estimation at every instant, which has been applied to a CSTR model. However, the covariance calculations encountered in the RNDDR formulation are similar to the EKF, namely, unconstrained propagation and correction involving the Kalman gain, which can affect the accuracy of the estimates. Recently, in order to overcome this disadvantage of the RNDDR, an unscented recursive nonlinear dynamic data reconciliation (URNDDR) technique[5] was proposed, which combined the merits of the unscented Kalman filter (UKF)[6] and the RNDDR to improve accuracy of the estimates. The application of the UKF to chemical processes were recently reported[7−8]. However, both the RNDDR and URNDDR could not reduce the affects of gross errors within the measurements.

More recently, particle filters[9] have attained significant interest with respect to state estimation, and an application of particle filters to dynamic data reconciliation was proposed[10], whose goal was to attain satisfactory state estimates and to detect the presence of gross errors. Though the particle filters approach could achieve more accuracy estimates, it usually takes much time and cannot take account of bounds and algebraic constrains for the state estimation.

Simultaneous data reconciliation and gross error detection can be addressed as a model identification and parameter estimation problem, since gross errors could be considered as parameters in the data reconciliation model, which could be estimated by using the reconciled values of the process variables[11]. Since simultaneous data reconciliation and gross error detection can be addressed as model identification and parameter estimation problem, support vector (SV) regression is introduced. The SV algorithm is a nonlinear generalization of the generalized portrait algorithm developed in Russia in the sixties of last century. As such, it is firmly grounded in the framework of statistical

Received December 26, 2008; in revised form March 14, 2009
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA040308), National Natural Science Foundation of China (60736021), and the National Creative Research Groups Science Foundation of China (60721062)
1. National Key Laboratory of Industrial Control Technology, Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310027, P. R. China
DOI: 10.1016/S1874-1029(08)60092-0

learning theory, or Vapnik-Chervonenkis (VC) theory, which has been developed over the last three decades by Vapnik[12] and others. According to statistical learning theory, minimizing empirical risk, which will lead to overfitting and thus bad generalization properties, is replaced by minimizing regularized risk with adding a capacity control term to the objective function[13]. Within the framework of statistical learning theory, or VC theory, the effects of gross errors could be considered as VC dimension, so the effects of gross errors could be eliminated from the estimation of real state variables. With minimizing the regularized risk instead of the maximum likelihood in the RNDDR, our approach could achieve not only recursive nonlinear dynamic data reconciliation but also gross error detection simultaneously. Meanwhile, we use filtered state estimates and measurements to estimate real states at right instant, so the predicted state estimates are not necessary in our approach, which overcomes the effects of the unaccurate covariance calculations in the linearized nonlinear dynamic system model.

In this paper, the merits of the RNDDR and SV regression are combined to obtain an SV regression approach for recursive simultaneous data reconciliation and gross error detection in nonlinear dynamic systems. Since this integration is achieved without sacrificing the recursive nature of the estimation procedure, a more accurate and efficient real-time recursive simultaneous data reconciliation and gross error detection for nonlinear dynamic processes is obtained. The execution of a comparative study between optimization approach RNDDR and our approach is considered in this paper, but probabilistic filtering methods are beyond the scope of this paper.

## 1 Recursive estimation techniques

In order to motivate the development of our approach, we first give a brief description of two recursive estimation techniques for nonlinear dynamic processes, the extended Kalman filters (EKF), and recursive nonlinear dynamic data reconciliation (RNDDR), which is based on Kalman filter.

Consider a process described by the following continuous-time nonlinear state space model with additive uncertainties (state noise) and discrete measurements sampled data at regular intervals:

$$\begin{aligned}\boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \int_{k\Delta t}^{(k+1)\Delta t} \boldsymbol{f}\left(\boldsymbol{x}\left(\tau\right), \boldsymbol{u}_k\right) \mathrm{d}\tau + \boldsymbol{w}_k \\ \boldsymbol{y}_{k+1} &= \boldsymbol{g}\left(\boldsymbol{x}_{k+1}\right) + \boldsymbol{v}_{k+1}\end{aligned} \tag{1}$$

In the above model, $\boldsymbol{x}_{k+1}$ is the $n \times 1$ vector of state variables, $\boldsymbol{y}_{k+1}$ is the $m \times 1$ vector of measurements, and $\boldsymbol{w}_k$ and $\boldsymbol{v}_{k+1}$ are mutually independent normally distributed random variables with covariance matrices $Q_k$ and $R_{k+1}$, respectively. The subscript $k$ represents time instant $t_k = k\Delta t$.

Assume that at time $t_k$ we have filtered state estimates denoted by $\hat{\boldsymbol{x}}_{k|k}$ which have been obtained using all the measurements up to time $t_k$. From these, the predicted state estimates $\hat{\boldsymbol{x}}_{k+1|k}$ at time $t_{k+1}$ are obtained as

$$\hat{\boldsymbol{x}}_{k+1|k} = \hat{\boldsymbol{x}}_{k|k} + \int_{k\Delta t}^{(k+1)\Delta t} \boldsymbol{f}\left(\boldsymbol{x}\left(\tau\right), \boldsymbol{u}_k\right) \mathrm{d}\tau + \boldsymbol{w}_k$$

$$\hat{\boldsymbol{x}}_{k|k} \equiv \hat{\boldsymbol{x}}\left(k\Delta t\right) \tag{2}$$

The predicted state estimates are corrected using the measurements $\boldsymbol{y}_{k+1}$, and by using the following linear update equation to obtain the filtered state estimates at time $t_{k+1}$.

$$\hat{\boldsymbol{x}}_{k+1|k+1} = \hat{\boldsymbol{x}}_{k+1|k} + K_{k+1}\boldsymbol{V}_{k+1} \tag{3}$$

where

$$\boldsymbol{V}_{k+1} = \boldsymbol{y}_{k+1} - \boldsymbol{g}\left(\hat{\boldsymbol{x}}_{k+1|k}\right) \tag{4}$$

and matrix $K_{k+1}$ is known as the Kalman gain. The computation of the Kalman gain for the EKF is described in the following section.

### 1.1 Extended Kalman filter (EKF)

Assume the covariance matrix of errors in the filtered state estimates at time $t_k$ is given by $P_{k|k}$. In EKF, the covariance matrix of errors in the predicted state estimates at time $t_{k+1}$ is approximated by linearizing the nonlinear state space model around $\hat{\boldsymbol{x}}_{k|k}$. The state space matrix for the linearized continuous time model is given by

$$A_k = \frac{\partial \boldsymbol{f}\left(\boldsymbol{x}, \boldsymbol{u}\right)}{\partial \boldsymbol{x}}\Big| \hat{\boldsymbol{x}}_{k|k}, \boldsymbol{u}_k \tag{5}$$

The covariance matrix of estimation errors in the predicted estimates is obtained as

$$P_{k+1|k} = \bar{A}_k P_{k|k} \bar{A}^{\mathrm{T}} + Q_k \tag{6}$$

where $\bar{A}_k = \exp\left(A_k \Delta t\right)$ is the state transition matrix for the equivalent linear discrete system. The Kalman gain matrix is computed using

$$K_{k+1} = P_{k+1|k} G_{k+1}^{\mathrm{T}} \left(G_{k+1} P_{k+1|k} G_{k+1}^{\mathrm{T}} + R_{k+1}\right)^{-1} \tag{7}$$

where $G_{k+1}$ is the linearized measurement model matrix given by $G_{k+1} = \frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}} | \hat{\boldsymbol{x}}_{k+1|k}$. The covariance matrix of errors in the updated state estimates is approximated using

$$P_{k+1|k+1} = \left(I - K_{k+1} G_{k+1}\right) P_{k+1|k} \tag{8}$$

There are several drawbacks associated with EKF, that is, EKF cannot ensure the estimated states to satisfy bounds and other algebraic constraints, and the covariance matrix of estimation errors and Kalman gain are calculated using an approximated linear model of the process which can have an adverse effect on the accuracy of the state estimates. Furthermore, EKF cannot eliminate the effects of gross errors within the measurements.

### 1.2 Recursive nonlinear dynamic data reconciliation

The update equation for KF as well as EKF can be obtained as the solution of an optimization problem[4]. So the RNDDR method was developed, which can take account of algebraic constraints and bound constraints to be satisfied by state estimates.

Consider the nonlinear dynamic system given by (1). The bounds and algebraic constraints of states are $\boldsymbol{x}_{\mathrm{L}} \leq \boldsymbol{x} \leq \boldsymbol{x}_{\mathrm{U}}$, and algebraic constraints of states are $h\left(\boldsymbol{x}, \boldsymbol{u}\right) \leq 0$, $e\left(\boldsymbol{x}, \boldsymbol{u}\right) = 0$. Let $\hat{\boldsymbol{x}}_{k|k}$ be the filtered estimates at time instant $k$ and the corresponding estimate error covariance matrix be $P_{k|k}$. In the RNDDR method, the predicted state and the covariance matrix of errors in the predicted estimates are obtained as in EKF using (2) and (6), and instead of using (3), the updated state estimates in the RNDDR method are obtained by solving the