Computers and Chemical Engineering 69 (2014) 26–38

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Application of stepwise regression for dynamic parameter estimation

Mordechai Shacham[a,*], Neima Brauner[b]

[a] Department of Chemical Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
[b] School of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel

ARTICLE INFO

Article history:
Received 25 December 2013
Received in revised form 22 June 2014
Accepted 25 June 2014
Available online 3 July 2014

Keywords:
Stepwise regression
Dynamic parameter estimation
Regression analysis

ABSTRACT

Dynamic parameter estimation in cases where it may be impossible to identify all the model parameters is considered. The objective is to obtain reliable estimates to the maximal number of physical parameters in a stable regression model where the modeling of the noise in the data is avoided. The modifications required in the stepwise regression algorithm to accommodate various nonlinear terms in the regression model are investigated and a new algorithm is presented. The algorithm considers the hierarchy among the parameters, the initial trends of the experimental data curves and the initial values of the state variables in order to establish a minimal initial set of parameters to be included in the model. Additional parameters are then added in a stepwise manner, while considering the hierarchy of the parameters and the associated reduction of the objective function value. The process continues as long as significant and physically feasible values for the parameters are obtained. The new method is demonstrated with several examples from the literature. Additional issues investigated include the proper combination of the simultaneous and sequential solution methods in the stepwise regression algorithm, the preferred method for the estimation of the derivatives and the effect of variable scaling.

1. Introduction

Reliable mechanistic kinetic models of chemical/biological processes are essential for the understanding, design, optimization and control of such processes. Such models are often described by systems of ordinary differential equations (ODEs). The models usually contain unknown physical parameters that need to be determined using a set of measurements (experimental data). Typically, the estimation of the parameter values is performed using a maximum likelihood approach, where the objective is to minimize the weighted squared error between the set of the measured data and the model predictions.

The various stages of the mechanistic model development and the estimation of the model parameters in dynamic systems were described in detail by Maria (2004). Sun et al. (2012) provide a recent review of the methods used for parameter estimation. They classify the optimization algorithms used as deterministic global, metaheuristic global (such as evolutionary algorithms, EA) and deterministic local (gradient based). They point out that deterministic global algorithms are very time consuming and usually cannot obtain satisfactory solutions in a reasonable time frame when applied to practical problems of realistic sizes. Consequently, they recommend the use of metaheuristics for identifying near optimal parameter values, followed by the use of gradient based methods in order to refine these values. Recently developed methods for finding global optimum in dynamic parameter estimation problems include, for example, the incremental identification procedure followed by parameter estimation (Michalik et al., 2009), decomposition of the problem by generating an artificial neural network model (ANN) that is consequently used to obtain an estimate of the parameters (Dua, 2011), and conversion of the dynamic system of equations into a set of algebraic constraints that enables the use of a deterministic global optimization approach for parameter identification (Esposito and Floudas, 2000).

Test problems for evaluation of the methods for their global convergence ability and for computational efficiency were proposed for example by Biegler et al. (1986), Floudas et al. (1999) and Moles et al. (2003). Many of those test problems, however, disregard some of the major difficulties involved in the tasks of the mechanistic model development and parameter identification.

Consider for example the test problem presented by Moles et al. (2003). This test problem involves the estimation of 36 kinetic parameters of a nonlinear biochemical model formed by 8 ODEs

* Corresponding author. Tel.: +972 8 6461481; fax: +972 8 6472916.
E-mail addresses: shacham@bgu.ac.il (M. Shacham), brauner@eng.tau.ac.il (N. Brauner).

http://dx.doi.org/10.1016/j.compchemeng.2014.06.013
0098-1354/

(ordinary differential equations) that describe the variation of the metabolite concentration with time. "Pseudo" experimental data for this test problem was generated by introducing a "nominal" set of parameter values into the model, integrating the model equations and recording the "experimental" values at 20 equally spaced time intervals. No measurement noise was added to the simulated experimental data. This test problem has proven to be quite challenging because of the high dimensionality and the large number of local solutions, however it represents oversimplification in many aspects related to practical problems. In this test problem there is no uncertainty regarding the adequacy of the mathematical model to represent the process (as the "process" data is actually generated by the model), and large amount of high precision "experimental" data, as well as a good set of references for initial guesses for the parameters (i.e., the parameter values that were used to generate the data) are available. Moreover, no attempt was made to check the validity of the parameter sets corresponding to the local (and global) solutions using statistical metrics (such as confidence intervals) or judging their feasibility from the physical point view.

Test problems which are more realistic in terms of the postulated mathematical model of the process and the available data are presented, for example, by Biegler et al. (1986), Maria (1989), Zamostny and Belohlav (1999), Kadam et al. (2004), Liu and Wang, 2009, Gennemark and Wedelin (2009), Brun et al. (2001) and Tsai et al. (2014). These problems are characterized by uncertainty regarding the suitability of the mathematical model to represent the data, availability of insufficient amount and/or low precision noisy experimental data and lack of sensible initial estimates for some (or all) of the parameter values. Several researchers (see for example, McLean and McAuley, 2012; McLean et al., 2012; Brun et al., 2001; Liu and Wang, 2009; Sin et al., 2010; Kravaris et al., 2013) demonstrate that in realistic parameter estimation problems it is impossible to identify all the parameters. Statistical metrics, such as confidence intervals, are used for determining the uncertainty and stability of the calculated parameter values. Due to the nonlinearity of the models, the parameter sensitivity information is obtained by linearization of the model in the vicinity of an initial estimate of the solution (McLean and McAuley, 2012) or at the optimum found (Sin et al., 2010). Parameter estimation in linear models usually involves removal of parameters (and their associated terms) if the value of zero is included within their confidence interval (see for example, Shacham and Brauner, 2003). Application of the same principle to the nonlinear models usually involved in dynamic parameter estimation problems may lead to extensive modification of the model upon removal of the terms associated with the unidentifiable parameters, which may result in unexpected consequences. In extreme cases the degenerated model may not represent the underlying physical phenomenon.

To avoid such a situation we have developed a new stepwise regression procedure with the objective to obtain reliable estimates to the maximal number physical parameters involved in the model whose values are unknown. These are often needed for further analysis and modeling of the pertinent physical phenomena. The procedure starts with a minimal set of parameters that can yield a feasible solution that is consistent with the initial trend of the state variables data, and continues with addition of parameters as long as significant and physically feasible values for the parameters are obtained. The optimal parameter values obtained in the earlier step are used as initial estimates for next step, to reduce the parameter search space and speed-up the computation. The details of this procedure are presented in Section 3 and its use is demonstrated by solving the examples provided by Zamostny and Belohlav (1999), Maria (1989) and Merchuk (2013).

2. Basic concepts

The ODE parameter estimation problem can be defined

$$\min_{\theta} \Phi_1(\theta) = \sum_{\mu=1}^{n}\sum_{i=1}^{m}(x_{\mu,i} - x^c_{\mu,i})^2 * \mathrm{W}_i \tag{1}$$

subject to

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, \boldsymbol{\theta}, t) \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2}$$

where $\mathbf{F}$ is a system of m ordinary differential equations, $\mathbf{x}$ is a vector of m dynamic variables, $\boldsymbol{\theta}$ is a vector of p parameters, $\mathbf{x}_\mu$ and $\mathbf{x}^c_\mu$ are the vectors of m observed and calculated values at the μth data point, respectively, and $\mathbf{W}$ is a vector of m weighting factors.

An alternative formulation which enables converting the differential parameter estimation problem into an algebraic parameter estimation problem is the following:

$$\min_{\theta} \Phi_2(\theta) = \sum_{\mu=1}^{n}\sum_{i=1}^{m}(\dot{x}_{\mu,i} - \dot{x}^c_{\mu,i})^2 * W'_i \tag{3}$$

subject to

$$\dot{\mathbf{x}}_\mu = \mathbf{F}(\mathbf{x}_\mu, \theta, t) \quad \mu = 1, 2, \ldots, n \tag{4}$$

where $\dot{x}_{\mu,i}$ is an estimate for the derivative: $\dot{x}_{\mu,i} = dx_i/dt$ at the μth data point. Using this formulation the matrix of the observed data points $\mathbf{x}_\mu$ replaces the vector of state variables $\mathbf{x}$ in the model $\mathbf{F}$.

With the first problem formulation, the solution is carried out using the sequential (or the feasible path) approach. With this approach, the minimization defined in Eq. (1) is carried out in an outer loop, while in the inner loop an integration routine is used to determine the state variable values at time intervals where experimental data are available. In this study, the integration is usually carried out by the MATLAB[1] *ode45* function, which is based on an explicit Runge–Kutta (4,5) formula, the Dormand–Prince (1980) pair. This algorithm monitors the integration-error estimate and controls the step size of the integration in order to keep the error below a specified threshold. The accuracy requested is that both the relative and absolute (maximal) errors be less than the truncation error tolerance. The default value of this tolerance is 1.0E−6. The integration is carried out in a piecewise manner from the point t_μ to $t_{\mu+1}$. At this point the μ+1th component of the objective function is calculated and added to the previous components of the objective function. This process is continued until reaching the last set of data points where $t_{\mu+1} = t_n$. For integration of stiff systems, the MATLAB library function *ode15s* is used, which is based on the backward difference formulas (BDF) method of Gear (1971).

For the outer loop minimization the Levenberg–Marquardt (LM) algorithm (Seber and Wild, 2003) as implemented in the MATLAB *nlinfit* function is used. This method belongs to the gradient-based category, where derivatives of the objective function and the constraints need to be evaluated. LM is a modification of the Newton method, where the second derivatives appearing in the Hessian matrix are neglected. The algorithm switches to the steepest descent method when the Hessian matrix becomes nearly singular.

Another solution technique involves the conversion of the differential parameter estimation problem into an algebraic parameter estimation problem. In order to use this technique, the derivatives of the state variables, $\dot{x}_{\mu,i}$ need to be estimated first. This

[1] MATLAB is a product of MathWorks, Inc., http://www.mathworks.com.