Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 48-8 (2015) 147–152

A Bayesian Method for Estimating Parameters in Stochastic Differential Equations

Hadiseh Karimi*. Kimberley B. McAuley**

Department of Chemical Engineering, Queen's University, Kingston, K7L3N6
**Canada (e-mail: hadiseh.karimi@queensu.ca).*
***Canada (Tel: 613-533-6000 ext.77562; e-mail: kim.mcauley@queensu.ca).*

Abstract: A Bayesian algorithm is developed for estimating parameters in nonlinear stochastic differential equation (SDE) models. The proposed algorithm uses prior information about parameters and builds on the approximate expectation maximization (AEM) algorithm (Karimi and McAuley, 2014a). A nonlinear continuous stirred tank reactor (CSTR) model is used to compare the effectiveness of the Bayesian algorithm to that of the AEM algorithm. For the CSTR example studied, the proposed method provides more accurate parameter estimates, especially for small data sets.

Keywords: Parameter estimation, maximum likelihood, stochastic modeling, stochastic systems, nonlinear equations, uncertain dynamic systems, differential equations, noise levels.

1. INTRODUCTION

Fundamental dynamic models are derived using material, energy and momentum balances. Stochastic terms are sometimes introduced on the right-hand sides of the resulting differential equations to account for disturbances and model mismatch (Jones et al., 1989). The resulting equations are called stochastic differential equations (SDEs). In this paper, we consider Multi-Input Multi-Output (MIMO) nonlinear SDE models of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \ , \tag{1.a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \, , \tag{1.b}$$

$$\mathbf{y}(t_{m\,r,j}) = \mathbf{g}(\mathbf{x}(t_{m\,r,j}), \mathbf{u}(t_{m\,r,j}), \boldsymbol{\theta}) + \boldsymbol{\varepsilon}(t_{m\,r,j}) \, , \tag{1.c}$$

where $\mathbf{x} \in R^X$ is the vector of state variables, t is time, $\mathbf{f} : R^X \times R^U \times R^P \to R^X$ is a vector of nonlinear functions, $\mathbf{u} \in R^U$ is the vector of input variables and $\boldsymbol{\theta} \in R^P$ is the vector of unknown model parameters. $\boldsymbol{\eta}(t) \in R^X$ is a continuous zero-mean stationary Gaussian white-noise process with covariance matrix $E\{\boldsymbol{\eta}(t_1)\boldsymbol{\eta}(t_2)\}=\mathbf{Q}\ \delta(t_2-t_1)$, where $\mathbf{Q}$ is the corresponding diagonal power spectral density function :

$$\mathbf{Q} = \begin{bmatrix} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_X \end{bmatrix} . \tag{2}$$

The matrix of power spectral density function sometimes referred to as the process disturbance intensity matrix (Varziri et al., 2008). $\delta(.)$ is the Dirac delta function and $\mathbf{x_0} \in R^X$ is a vector of initial conditions for the state variables. Some of these initial conditions may be known to the modeler and others may be unknown values that require estimation along with the model parameters. $\mathbf{y} \in R^Y$ is the vector of measured output variables. The times at which measurements are available for the rth response (r=1…Y) are denoted by $t_{m\,r,j}$ (j = 1…N_r) where N_r is the number of measurements for the rth response. $\mathbf{g} \in R^Y$ is a vector of nonlinear mappings and $\boldsymbol{\varepsilon} \in R^Y$ is a vector of zero-mean random variables. If the measurement errors are independent, the corresponding covariance matrix has the following form:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_Y^2 \end{bmatrix} . \tag{3}$$

Consider the vector $\mathbf{Y_m}$ that contains the stacked measured value $\mathbf{Y_m} = [y_1(t_{m\,1,1}) \ldots y_1(t_{m\,1,N_1}) \ldots y_Y(t_{m\,Y,1}) \ldots y_Y(t_{m\,Y,N_Y})]^{\mathrm{T}}$ and $\mathbf{X_m} = [x_1(t_{m\,1,1}) \ldots x_1(t_{m\,1,N_1}) \ldots x_Y(t_{m\,Y,1}) \ldots x_Y(t_{m\,Y,N_Y})]^{\mathrm{T}}$ which contains the stacked values of the state variables at the measurement times. $\mathbf{U_m}$ and $\boldsymbol{\varepsilon_m}$ are corresponding vectors for the input variables and random errors, respectively so that:

$$\mathbf{Y_m} = \mathbf{G}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon_m} \, . \tag{4}$$

where $\mathbf{G}$ is $\mathbf{G}=[\mathbf{g}, \ldots, \mathbf{g}]^{\mathrm{T}}{}_{1\times NY}$. The index m for a variable indicates that the values of that variable are taken at measurement times. The existence of a solution of an SDE is ensured when globally Lipschitz, linear growth and boundedness conditions are satisfied (Liptser and Bishwal, 2000). Since $\boldsymbol{\eta}(t)$ does not have a simple mathematical interpretation, SDEs are often written in the differential form (Liptser and Shiryaev, 2000):

$$\mathrm{d}\,\mathbf{x} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})dt + \mathbf{Q}d\mathbf{W} \, , \tag{1.d}$$

10.1016/j.ifacol.2015.08.172

whrere $\mathbf{W}(t)$ is a Wiener process. Modelers often have knowledge about the accuracy of their measurements (*i.e.*, the diagonal elements of $\mathbf{\Sigma}$), but do not have knowledge about the values of the diagonal elements of $\mathbf{Q}$, (i.e., $\mathbf{Q_d}=[Q_1,\ldots,Q_X]^T$). Let $\boldsymbol{\zeta} = [\boldsymbol{\theta}^T, \mathbf{x}_{0u}^T, \mathbf{Q_d}^T]^T$ be the vector of unknown parameters in the SDE model where $\mathbf{x}_{0u}$ is a vector of the unknown initial conditions. SDE models are used for simulation, design and optimization of chemical processes and for model predictive control (McLean and McAuley, 2012). Therefore, accurate and reliable parameter estimation techniques for SDE models are beneficial for chemical engineers. Maximum likelihood estimation (MLE) methods are commonly used to estimate parameters in SDE models because of their asymptotic efficiency and consistency (Casella and Berger, 1990).

A challenge in estimating parameters in SDE models for chemical engineering systems is that experiments and measurements are often limited due to cost or difficulties in measuring certain variables. Performing further experiments is expensive and may not be feasible (McLean and McAuley, 2012). As a result, the number of data values for parameter estimation may be limited and some of the states are often not measured. While MLE methods provide satisfactory solutions for parameter estimation in many SDE models, they provide noticeably biased parameter estimates when only limited data are available for parameter estimation (Casella and Berger, 1990; Ninness and Henriksen, 2010).

In chemical engineering applications, prior information about some of the parameters is often known to the modeler (e.g., reasonable initial guesses and physically realistic ranges for parameter values). Box and Draper (1964) introduced the use of Bayesian methods for estimating parameters in chemical engineering models so that prior knowledge about parameter values could be accounted for. One benefit of Bayesian parameter estimation methods is that they can provide improved parameter estimates, especially when available datasets are small (Robert and Casella, 1999). In general, in Bayesian methods, the probability density function of the parameters given the measured data $p(\boldsymbol{\zeta} \mid \mathbf{Y_m})$ is maximized to estimate the unknown parameters. This joint probability function is referred to as the posterior density function (Jang and Gopaluni, 2011). The posterior density function can be obtained from:

$$p(\boldsymbol{\zeta} \mid \mathbf{Y_m}) = \frac{p(\mathbf{Y_m} \mid \boldsymbol{\zeta})\, p(\boldsymbol{\zeta})}{p(\mathbf{Y_m})} . \tag{5}$$

The numerator on the right-hand side is the product of the probability density function of the measurements given parameters $p(\mathbf{Y_m} \mid \boldsymbol{\zeta})$ and the prior distribution of the parameters $p(\boldsymbol{\zeta})$, which contains knowledge about the possible values of $\boldsymbol{\zeta}$. The likelihood function of the parameters given the measurements is defined as

$$L(\boldsymbol{\zeta} \mid \mathbf{Y_m}) = p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) . \tag{6}$$

The denominator in (5), which ensures that the posterior integrates to unity, does not depend on the parameter values. The prior probability $p(\boldsymbol{\zeta})$ is important when there is limited data available to provide reliable estimates for some of the model parameters. When a large quantity of informative data is available, the posterior probability will be dominated by the likelihood function In nonlinear models with unmeasured states, evaluation of the posterior density function is a major challenge requiring calculation of complicated integrals of probability density functions (Jang and Gopaluni, 2011; Ljung, 1999). Computationally intensive Markov Chain Monte Carlo (MCMC) algorithms (also referred to as particle filtering), which require very few assumptions about the posterior density function, have been used to compute these integrals (Coleman and Block, 2006; Jang and Gopaluni, 2011; Robert and Casella, 1999). MCMC methods are used to approximate the posterior densities in SDE models (Geweke and Tanizaki, 2001; Jang and Gopaluni, 2011; Ninness and Henriksen, 2010), in mixed models (Gelman, 2006) and in ordinary differential equation models (Coleman and Block, 2006). MCMC methods are particularly computationally demanding when the number of states and parameters is large (Gopaluni, 2010). Benefits and drawbacks of MCMC methods are summarized by Chen et al. (2004).

In this article a computationally efficient algorithm is proposed for estimating parameters and states in nonlinear SDE models when the modeler has some prior knowledge about some of the parameters. This algorithm is developed using a Bayesian approach. Recently, we developed three approximate MLE algorithms for estimating parameters in nonlinear SDE models (Karimi and McAuley, 2013; Karimi and McAuley, 2014a; Karimi and McAuley, 2014b). These MLE-based methods, which do not require prior knowledge about parameters, are computationally efficient, but can provide poor parameter estimates when data sets are small. Here, we develop an approximate Bayesian expectation maximization (ABEM) algorithm that builds on our previous approximate expectation maximization (AEM) method (Karimi and McAuley, 2014a). The inclusion of prior information about parameters in the resulting objective function leads to improved parameter estimates, especially when data are sparse or noisy. First, an analytical expression for the posterior density function is derived and used to develop a suitable objective function for parameter estimation. The proposed algorithm is then tested using a CSTR model and results are compared with those from the AEM method. It is shown that the proposed ABEM method provides more accurate parameter estimates for the example studied.

2. DEVELOPMENT OF THE APPROXIMATE BAYESIAN EXPECTATION MAXIMIZATION ALGORITHM

In Bayesian approaches, the posterior density function $p(\boldsymbol{\zeta} \mid \mathbf{Y}_m)$ is maximized to obtain the parameter estimates. Maximizing $p(\boldsymbol{\zeta} \mid \mathbf{Y}_m)$ is equal to minimizing -ln $p(\boldsymbol{\zeta} \mid \mathbf{Y}_m)$. When developing the AEM methodology, we showed that (Karimi and McAuley, 2014a):