Bayesian Method for Identification of Constrained Nonlinear Processes with Missing Output Data

Jing Deng and Biao Huang

Abstract— A methodology for the identification of nonlinear models using constrained particle filters under the scheme of the expectation-maximization (EM) algorithm is presented in this paper. Missing or irregularly sampled observations are commonplace in the chemical industry. In order to circumvent the difficulties rendered by largely incomplete data set, an improved EM based algorithm, which uses the expected value of the log-likelihood function including the missing observations, is developed. Constrained particle filters are adopted to solve the expected log-likelihood function in the EM algorithm. The efficiency of the proposed method in handling missing data is illustrated through numerical examples and validated through experiments.

Keywords: Parameter estimation, Nonlinear state space model, Missing outputs, Expectation-Maximization method, Constrained particle filters

I. INTRODUCTION

Over the past few decades, the research of parameter estimation has witnessed rapid progress as it plays a key role in the development of mathematical models (which describe the process behavior). Achieving a fairly accurate estimate of the model parameters is imperative to satisfactorily capture the system dynamics. Aspects of system identification have been discussed extensively in literatures [1], [2]. However, studies of parameter estimation in the presence of missing observations have been less reported. Missing observations or irregularly sampled outputs are commonly experienced in real industry process. Lack of complete process knowledge poses a big challenge to researchers and engineers in performing model parameters estimation. Some common approaches in dealing with missing data have been presented and summarized in Khatibisepehr and Huang (2008) [3], such as case-wise deletion, mean substitution, regression imputation, multiple imputation, etc. Nevertheless, as pointed out in Khatibisepehr and Huang (2008) [3], the variances of the values may be considerably changed if applying case-wise deletion, mean substitution, regression substitution, etc. for missing data treatment.

When dealing with a dynamic system with missing data, the methods that have been discussed would fail because of their incapability of handling the dependency of the current output or state on the previous ones. While it is beyond the scope of this paper to thoroughly cover all types of dynamic models, our special attention is paid to nonlinear state space models. In fact, quite limited work of identification of nonlinear dynamic models subject to missing data has been done.

The expectation-maximization (EM) algorithm [5] is known as one of the techniques that can elegantly handle missing outputs. Having long been recognized for its attractive statistical properties as well as the convenience to implement, some researches have reported its application in parameter estimation with missing observations. Shumway and Stoffer (1982) [6] use the EM algorithm in combination with the conventional Kalman smoothed estimators $x_t^n = E(x_t|y_1, y_2, \ldots, y_n)$ to perform the smoothing and forecasting for time series with missing observations. However, they only consider the case without manipulated inputs. In recent years, particle filters, which are also known as sequential Monte Carlo (SMC) [7] methods, have been utilized in combination with the EM algorithm [8], [9]. A maximum likelihood method of identifying parameters of a nonlinear process model with missing outputs is reported in Gopaluni (2008) [8], where the log-likelihood in the EM algorithm is approximated by particle filters and smoothers. Following the point-wise state estimation technique, the smoothed density functions of each state has to be calculated at every iteration in the EM algorithm, which introduces a large amount of computing burden. Based on [8], we propose a constrained Bayesian method with only filtering under the expectation-maximization (EM) framework to deal with the parameter estimation problem with missing data, resulting a more efficient scheme for nonlinear system identification.

Though it is recognized that nonlinear systems subject to constraints are commonly encountered in practice, the majority of existing Bayesian estimation methods are implemented by ignoring such constraints. Recently, some researchers have presented related work dealing with process constraints using particle filters [14], [15]. By including certain process constraints into the particle filtering procedure and performing cautious resampling strategy, particles are ensured to efficiently represent the true density. Meanwhile, the diversity of particles can be preserved to certain extent.

The remainder of this paper is organized as follows: Section 2 states the identification problem of nonlinear system. Section 3 begins with a revisit of the EM algorithm and then the derivation of the expression for Q function with missing data is given. Section 4 provides a brief description of constrained PF and the details of approximating the Q

J. Deng is with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada, T6G-2G6 `jdeng4@ualberta.ca`

B. Huang is with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada, T6G-2G6 `biao.huang@ualberta.ca`

function using constrained PF are presented. Numerical simulations are illustrated in Section 5 which aim at demonstrating the effectiveness of the proposed method in nonlinear system identification with missing output data. Comparative studies are also performed with the proposed method and a different missing data treatment method. Section 6 draws the conclusion based on the results obtained in this paper.

II. PROBLEM STATEMENT

Consider the state space model given by

$$x_t = f(x_{t-1}, u_{t-1}) + \omega_t \tag{1}$$

$$y_t = h(x_t) + v_t \tag{2}$$

where the system parameters are Θ. x_t, u_t, y_t, ω_t and v_t are state, measured input, measured output, process noise and measurement noise, respectively; ω_t and v_t are independent and identically distributed Gaussian noises with covariance matrices Q, and R respectively. The input sequence $\{u_1, \ldots, u_t\}$ is known.

Let X denote the sequence of hidden states $\{x_1, \ldots x_T\}$. Suppose that the outputs are available at time $\{t_1, \ldots, t_\alpha\}$ while missing at time $\{m_1, \ldots, m_\beta\}$. $Y_o = \{y_{t_1}, \ldots, y_{t_\alpha}\}$ and $Y_m = \{y_{m_1}, \ldots, y_{m_\beta}\}$ stand for the corresponding observed outputs and missing outputs. It is assumed that the data is missing completely at random (MCAR). In other words, the probability that the data is missing does not depend on any part of the observed data. The process and measurement model structures in Equation (2) are assumed to be known a priori. Thus, determination of the model structure is beyond the scope of this paper. Here the parameters in the state space model, Θ, are of interest. In the following section, we show how to formulate the parameter estimation problem under the framework of the EM algorithm.

III. EXPECTATION-MAXIMIZATION ALGORITHM

A. EM algorithm revisit

Expectation-maximization (EM) algorithm [5] is a well-known maximum likelihood based method, which operates between two steps, the expectation step and maximization step. The basic principle behind the EM algorithm is that instead of performing a complicated optimization, one augments the observed data Y_{obs} with latent data Z to perform a series of optimizations. Hence both the complete-data log-likelihood $\log[(Y_{obs}, Z|\Theta)]$ and the conditional predictive distribution $p(Z|Y_{obs}, \Theta)$ can be calculated. Consisting of two steps, namely the expectation step (E-step) and the maximization step (M-step), the EM algorithm proceeds as follows.

Specifically, let Θ^k be the current best approximation to the mode of the observed posterior or the best estimated

parameters using all available data. The E-step is to compute the Q function which is formally defined by

$$Q(\Theta|\Theta^k) = E_{Z|Y_{obs}, \Theta^k}\{log[p(Y_{obs}, Z|\Theta)]\}$$
$$= \int_Z log[p(Y_{obs}, Z|\Theta)]p(Z|Y_{obs}, \Theta^k)dZ \tag{3}$$

and the M-step is to maximize the Q function with respect to Θ to obtain

$$\Theta^{k+1} = \arg\max_\Theta Q(\Theta|\Theta^k) \tag{4}$$

B. Formulation of the parameter estimation based on the EM algorithm

Consider the state-space model described above. The observed output data Y_{obs} are $Y_o = \{y_{t_1}, \ldots, y_{t_\alpha}\}$ while the hidden states $X = \{x_1, \ldots, x_T\}$ and the missing outputs $Y_m = \{y_{m_1}, \ldots, y_{m_\beta}\}$ can be viewed as the latent data Z. Let $p(x_{1:T}, y_{1:T}|\Theta)$ denote the complete likelihood function of the hidden states and observations. The Q function is defined as the expectation of the log-likelihood function $log[p(x_{1:T}, y_{1:T}|\Theta)]$ which is an integral given by

$$Q(\Theta|\Theta^k) = E_{Z|Y_{obs}, \Theta^k}\{log[p(Y_{obs}, Z|\Theta)]\}$$
$$= E_{X, Y_m|Y_o, \Theta^k}\{log[p(Y_o, X, Y_m|\Theta)]\}$$
$$= \int_{X, Y_m} log[p(Y_o, X, Y_m|\Theta)]p(X, Ym|Y_o, \Theta^k)dXdY_m$$
$$= \int_{X, Y_m} log[p(x_{1:T}, y_{1:T}|\Theta)]p(x_{1:T}, y_{m_1:m_\beta}|y_{t_1:t_\alpha}, \Theta^k)$$
$$\cdot dx_{1:T}dy_{m_1:m_\beta} \tag{5}$$

Following the approach of [8], the Q function can be derived. In Equation (5), the first term which is the joint density function of states and outputs can be decomposed using the Markov property as

$$p(x_{1:T}, y_{1:T}|\Theta) = p(x_{1:T}|\Theta)p(y_{1:T}|x_{1:T}, \Theta) \tag{6}$$

which can be further rewritten as

$$p(x_{1:T}, y_{1:T}|\Theta) = p(x_1|\Theta)\prod_{t=2}^T p(x_t|x_{t-1}, \Theta)$$
$$\cdot \prod_{t=1}^T p(y_t|x_t, \Theta) \tag{7}$$

Furthermore, substituting Equation (7) into Equation (5), the Q-function with missing outputs is given by

$$Q(\Theta|\Theta^k)$$
$$= \int_{X, Y_m} log[p(x_{1:T}, y_{1:T}|\Theta)]p(x_{1:T}, y_{m_1:m_\beta}|y_{t_1:t_\alpha}, \Theta^k)$$
$$\cdot dx_{1:T}dy_{m_1:m_\beta}$$
$$= \int_X log[p(x_1|\Theta)]p(x_{1:T}|y_{t_1:t_\alpha}, \Theta^k)dx_{1:T}$$
$$+ \sum_{t=2}^T \int_X log[p(x_t|x_{t-1}, \Theta)]p(x_{1:T}|y_{t_1:t_\alpha}, \Theta^k)dx_{1:T}$$
$$+ \sum_{t=t_1}^{t_\alpha} \int_X log[p(y_t|x_t, \Theta)]p(x_{1:T}|y_{t_1:t_\alpha}, \Theta^k)dx_{1:T}$$
$$+ \sum_{t=m_1}^{m_\beta} \int_{X, Y_m} log[p(y_t|x_t, \Theta)]p(x_{1:T}, y_{m_1:m_\beta}|y_{t_1:t_\alpha}, \Theta^k)$$
$$\cdot dx_{1:T}dy_{m_1:m_\beta} \tag{8}$$

The detailed derivation of the Q-function in (8) can be found in [8]. Estimation of $p(x_{1:T}|y_{t_1:t_\alpha}, \Theta^k)$ is a batch state estimation problem with all available observations. The speed of the execution with the iterative EM algorithm is very slow. The method proposed in [8] attempted to simplify Equation (8) by further marginalization of state and missing observations, but the end result is still a smoothing problem. The computation cost remains very high. Thus in this paper we will solve the problem by recursive state filtering while taking into account of the system constraints such that $p(x_{1:T}|y_{t_1:t_\alpha}, \Theta^k)$ is recursively approximated by $p(x_t|y_{t_1:t_\beta}, \Theta^k)$ for $t = 1 : T$, where $t_\beta \leq t$. The lost performance from smoothing to filtering is, to some extent, compensated by constraining the solution space, as demonstrated by simulation examples. This proposed solution can reduce the computation cost while keeping comparable accuracy in parameter estimation as that of the existing method, and thus make the solution feasible in real-time applications.

IV. CONSTRAINED BAYESIAN APPROACH

A. Particle filters revisit

The basic idea of particle filters is to represent the desired posterior density function by a series of particles with associated weights, i.e. $\{x_t^i, w_t^i\}_{i=1}^N$. Then the density function of the states given the current estimation of parameters Θ^k can be discretely approximated as

$$p(x_t|y_{t_1:t_\beta}, \Theta^k) \approx \sum_{i=1}^N \omega_t^i \delta(x_t - x_t^i) \tag{9}$$

where $\delta(\cdot)$ is the Dirac delta function, $t_\beta \leq t$; N is the number of particles; ω_t^i is the normalized weight associated with the ith particle such that $\sum_{i=1}^N \omega_t^i = 1$. Suppose that at time $t - 1$, a set of particles $\{x_{1:t-1}^i\}_{i=1}^N$ are available and we want to obtain N particles which represent the hidden state for time t. Since it is usually difficult to directly draw samples from the true posterior density $p(x_t|y_{t_1:t_\beta}, \Theta^k)$, the principle of importance sampling [11] is adopted. The idea is to use a so called importance density $q(\cdot)$ from which one can easily draw samples $x_t^i, i = 1, \ldots, N$. It has been shown that, as long as the support region of the posterior density belong to that of the importance density, the particle approximation is unbiased [12]. Owing to its simplicity to implement, the density defined by the state equation is adopted here for importance sampling, i.e.

$$q(x_t|y_{t_1:t_\beta}, \Theta^k) = p(x_t|x_{t-1}, \Theta^k) \tag{10}$$

With this choice, the unnormalized weight for each particle can be derived as [10]

$$\tilde{\omega}_t^i \propto \omega_{t-1}^i p(y_t|x_t^i, \Theta^k) \tag{11}$$

and the normalized weight can be calculated as

$$\omega_t^i = \frac{\tilde{\omega}_t^i}{\sum_{i=1}^N \tilde{\omega}_t^i} \tag{12}$$

For time instants $t = m_1, \ldots, m_\beta$, when the outputs are not available, draw particles from the importance density $p(x_t|x_{t-1}, \Theta^k)$ and keep the weights unchanged, i.e.

$$\omega_t^i = \omega_{t-1}^i \tag{13}$$

To avoid the degeneracy problem [10], the importance sampling step is usually followed by a resampling procedure. The idea is to discard the particles with small weights and concentrate on those with large weights. After resampling, each particle's weight will be reset to $\omega_t^i = \frac{1}{N}$.

B. Constrained particle filters approximation and cautious resampling

The problem brought by brutal force resampling is that it reduces the diversity among particles. One solution is to resample the particles only when it is necessary instead of performing it at each step. To be specific, N_{eff} is introduced to represent the effective particle number [13]

$$N_{eff} = \frac{1}{\sum_{i=1}^N (\omega_t^i)^2} \tag{14}$$

where ω_t^i is the normalized weight obtained by Equation (12). It can be implied that, as the variance of the weights grows very large, the effective sample size decreases to a small number which indicates a severe degeneracy problem. In practice, one uses resampling to eliminate useless particles only when a severe degeneracy problem occurs, say, N_{eff} falls below the threshold N_{thred}.

To further ensure the efficiency when generating particles, constraints are incorporated in the importance sampling step. A detailed discussion of constrained sequential monte carlo (SMC) algorithm can be found in Lang et al. (2007) [14]. At each importance sampling step, only those particles that satisfy the constraints can be accepted, which increases the proportion of useful particles.

Given the current estimation of parameters, the constrained particle filter algorithm is summarized as follows:

Step 1. Initialization. Draw initial N particles $\{x_0^i\}_{i=1}^N$ from the prior density $p(x_0|\Theta^k)$ and set each particle's weight to $\frac{1}{N}$. Set t=1.

Step 2. Importance sampling. Generate predicted particles $\{x_t^i\}_{i=1}^N$ from the importance density $p(x_t|x_{t-1}, \Theta^k)$ when the constraints are satisfied.

Step 3. Assigning weights. Assign the weight to each particle using Equation (11) and (12) when y_t is available. Otherwise, calculate the weights according to Equation (13).

Step 4. Resampling. Compute the number of effective particles using Equation (14). If N_{eff} is less than the threshold N_{thred}, then perform resampling and replace the predicted particles in Step 2 with resampled particles. Reset the weights of resampled particles uniformly as $\omega_t^i = \frac{1}{N}$. Otherwise, go to Step 5.

Step 5. Set $t = t + 1$ and repeat Step 2 to Step 4 for $t \leq T$.

In Equation (8), the density function $p(x_{1:T}|y_{t_1:t_\alpha}, \Theta^k)$ and $p(x_{1:T}, y_{m_1:m_\beta}|y_{t_1:t_\alpha}, \Theta^k)$ in the Q function can be

approximated using particle filters as

$$p(x_{1:T}|y_{t_1:t_\alpha}, \Theta^k) \approx \sum_{i=1}^{N} \omega^i \delta(x_{1:T} - x_{1:T}^i) \quad (15)$$

$$p(x_{1:T}, y_{m_1:m_\beta}|y_{t_1:t_\alpha}, \Theta^k) \approx$$
$$\sum_{i=1}^{N} \omega^i \delta(x_{1:T} - x_{1:T}^i) \delta(y_{m_1:m_\beta} - y_{m_1:m_\beta}^i) \quad (16)$$

In Equation (16), for the missing outputs part where the observations are not available at time $\{m_1, \ldots, m_\beta\}$, each missing output y_t is replaced by the sample drawn from the density function $p(y_t|x_t, \Theta^k), t = m_1, \ldots, m_\beta$.

Finally, the Q function in Equation (8) can be approximated as follows

$$Q(\Theta|\Theta^k) \approx \sum_{i=1}^{N} \omega^i log[p(x_1^i|\Theta)]$$
$$+ \sum_{t=2}^{T} \sum_{i=1}^{N} \omega_t^i log[p(x_t^i|x_{t-1}^i, \Theta)]$$
$$+ \sum_{t=t_1}^{t_\alpha} \sum_{i=1}^{N} \omega_t^i log[p(y_t|x_t^i, \Theta)]$$
$$+ \sum_{t=m_1}^{m_\beta} \sum_{i=1}^{N} \omega_t^i log[p(y_t|x_t^i, \Theta)] \quad (17)$$

With the approximated Q function, the EM algorithm can hence be implemented. In the expectation step, the Q function is evaluated according to Equation (17) with the current estimated parameters Θ^k. In the next maximization step, the new parameters Θ^{k+1} are obtained by maximizing the Q function.

To Maximize the Q function over parameters Θ, derivative operation is performed with respect to each parameter. Therefore, optimal system parameters at each iteration can be calculated by equating the derivatives to zero, i.e. $\frac{\partial Q}{\partial \theta_i} = 0$, where θ_i is the ith system parameter.

The EM algorithm is summarized as follows:

Step 1. Initialization. Start with the initial parameters Θ^0 and set t=0.

Step 2. Expectation. At time t, calculate the approximate Q function using Equation (17), given the current estimation of the system parameters Θ^k.

Step 3. Maximization. Maximizing the approximated Q function and get the new parameters Θ^{k+1}. Set k=k+1.

Step 4. Repeat Step 2 and Step 3 until the converge condition is satisfied, i.e. the change of the estimated parameters between two iterations is less than the tolerance.

V. SIMULATION EXAMPLES

In this section, the proposed approach is evaluated through both numerical simulations. Its efficiency in handling missing outputs with less computational cost will be demonstrated. All the simulations were run on a 3.00 GHz CPU with 4 GB RAM PC using MATLAB 2009a.

A. A Numerical Simulation Example

A nonlinear process considered in Goodwin et al. (2005) [16] and Gopaluni (2008) [8] is utilized here to demonstrate the efficiency of the proposed parameter estimation method. The process is described by the following equations:

$$x_t = ax_{t-1} + bu_{t-1} + \omega_t \quad (18)$$
$$y_t = ccos(x_t) + v_t \quad (19)$$

where $\omega_t \sim N(0, Q), v_t \sim N(0, R)$, u is a random binary input signal with unity variance, and $a = 0.9$, $b = c = 1$, $Q = R = 0.01$. State constraints are commonly experienced in practice. For example, the state of a partial pressure in a gas-phase reaction is subject to a non-negative constraint [17]. In a batch reactor system [18], the state of model fraction must be within the constraints between 0 to 1. In this example, the system states are subject to the constraint such that $-6 < x_t^i < 6$ from the prior system knowledge. T=100 measurements are collected from the simulation experiment. To test the algorithm's capability in handling the missing data, different portions of the output data are randomly removed from the model training data set. With the same condition as in Gopaluni (2008) [8], the estimation starts from the initial parameters $a_0 = b_0 = c_0 = 0.5$. N=150 particles are used for the particle filter approximation. The constraint is set to the particles $\{x_t^i\}_{i=1}^{N}$ which represent the state x_t such that $-6 \leq x_t^i \leq 6$ for all t. The proposed Bayesian parameter estimation method with the EM algorithm is performed. In the expectation step, the Q function is calculated according to Equation (17), where

$$log[p(x_t|x_{t-1}^i, \Theta)] = log[\frac{1}{\sqrt{2\pi Q_x}} \exp[-\frac{1}{2} \frac{(x_t^i - ax_{t-1}^i - bu_{t-1})^2}{Q_x}] \quad (20)$$

For $t = t_1 : t_\alpha$,

$$log[p(y_t|x_t^i, \Theta)] = log[\frac{1}{\sqrt{2\pi Q_y}} \exp[-\frac{1}{2} \frac{(y_t - c \cos x_t^i)^2}{Q_y}] \quad (21)$$

For $t = m_1 : m_\beta$,

$$log[p(y_t|x_t^i, \Theta)] = log[\frac{1}{\sqrt{2\pi Q_y}} \exp[-\frac{1}{2} \frac{(y_t^i - c \cos x_t^i)^2}{Q_y}] \quad (22)$$

where $y_t^i = c^{old} \cos x_t^i$.

By taking derivative over the Q function and equating it to zero, each individual component of the parameters is hence calculated as

$$a^{new} = \frac{\sum_{t=2}^{T} \sum_{i=1}^{N} \omega_t^i (x_t^i x_{t-1}^i - b^{old} x_{t-1}^i u_{t-1})}{\sum_{t=2}^{T} \sum_{i=1}^{N} \omega_t^i (x_{t-1}^i)^2} \quad (23)$$

$$b^{new} = \frac{\sum_{t=2}^{T} \sum_{i=1}^{N} \omega_t^i (x_t^i u_{t-1} - a^{old} x_{t-1}^i u_{t-1})}{\sum_{t=2}^{T} \sum_{i=1}^{N} \omega_t^i u_{t-1}^2} \quad (24)$$

$$c^{new} = \frac{\sum_{t=t_1}^{t_\alpha} \sum_{i=1}^{N} \omega_t^i y_t \cos x_t^i + \sum_{t=m_1}^{m_\beta} \sum_{i=1}^{N} \omega_t^i c^{old} (\cos x_t^i)^2}{\sum_{t=1}^{T} \sum_{i=1}^{N} \omega_t^i (\cos x_t^i)^2} \quad (25)$$

The trajectories of the estimated parameters are shown in Figure 1, Figure 2 and Figure 3. The estimated parameter values after 40 iterations are given in Table I.

TABLE I
ESTIMATED PARAMETERS AFTER 40 ITERATIONS

Proportion of missing output	a	b	c
10%	0.8925	1.0224	0.9960
25%	0.8903	0.9967	1.0180
50%	0.8878	0.9801	1.0021

TABLE II
ESTIMATED PARAMETERS AFTER 40 ITERATIONS WHEN 25% OUTPUTS ARE MISSING

	True parameters	Proposed method	Comparison method in [8]
a	0.9000	0.8965	0.9047
b	1.0000	0.9893	1.0073
c	1.0000	1.0177	0.9817
CPU time (s)	N/A	2.585×10	1.304×10^3

It can be seen from the above estimation result that, all parameters converge to the neighborhood of the true values after certain iterations. As the proportion of missing output increases, more iterations are needed for the EM algorithm to achieve the convergency. In Figure 1 where 10% observations are missing, estimated parameters barely change after about 15 iterations. However, as it is shown in Figure 3, nearly 25 iterations are needed for all parameters to achieve convergence when 50% observations are missing.

For comparison, the method proposed in Gopaluni (2008) [8] which uses a point-wise density function with smoothing in the approximation procedure is also applied to the same data set where 25% observations are missing. The estimation results using two methods are given in Table II. It can be seen from Table II that, the estimated parameters converge to the neighborhood of the true values using both methods. However, the approached method is much more efficient in terms of computation time.

B. Application on the continuous Stirred Tank Reactor

In this section, the popular continuous stirred tank reactor (CSTR), which has been accepted as a benchmark for process modeling is investigated. The governing model is given below [19]

$$\frac{dC_A(t)}{dt} = \frac{q(t)}{V}(C_{A0}(t) - C_A(t)) - k_0 C_{(A)}(t) exp(\frac{-E}{RT(t)}) \quad (26)$$

$$\frac{dT(t)}{dt} = \frac{q(t)}{V}(T_0(t) - T(t)) - \frac{(-\triangle H)k_0 C_A(t)}{\rho C_p} exp(\frac{-E}{RT(t)})$$

$$+ \frac{\rho_c C_{pc}}{\rho C_p V}q_c(t)\{1 - exp(\frac{-hA}{q_c(t)\rho C_p})\}(T_{c0}(t) - T_t) \quad (27)$$

The explanations for the components in the above equations are given in Table III along with their corresponding steady state values.

The concentration of the reactant C_A and the temperature in the reactor T are the two states of the process. The inlet flow rate q is considered as the process input. In the simulation experiment, a multiple level random input signal is designed. It is assumed that the process and measurement noise distributions are fairly small compared to the input variations. The input and output data of the process are given in Figure 4. The parameters being estimated are $\theta_1 = (-\triangle H)/\rho$ and $\theta_2 = (\rho_c C_p Cq_c)/(\rho C_p V)$. The estimation starts from the initial guess such that $\theta_1 = 0.5, \theta_2 = 1 \times 10^2$. N=150 particles are used for the particle filter approximation. The nonnegative constraint is set to the particles $\{x_t^i\}_{i=1}^N$ which

Fig. 1. Estimated parameters trajectories when 10% observations are missing. Blue solid line represents the trajectory of a; red star line represents the trajectory of b; green dotted line is the trajectory of c.
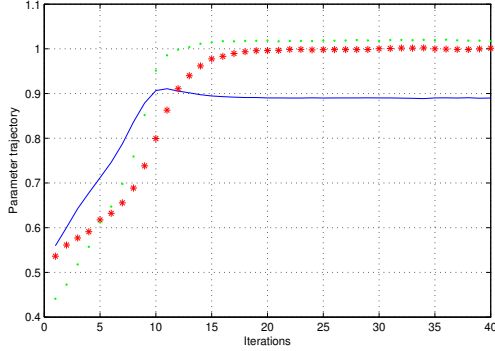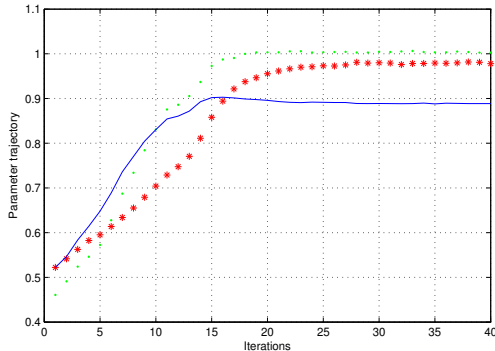

Fig. 2. Estimated parameters trajectories when 25% observations are missing. Blue solid line represents the trajectory of a; red star line represents the trajectory of b; green dotted line is the trajectory of c.

Fig. 3. Estimated parameters trajectories when 50% observations are missing. Blue solid line represents the trajectory of a; red star line represents the trajectory of b; green dotted line is the trajectory of c.

| Parameters	Steady state value
Cooling liquid flow rate, q_c	input
Production concentration of reactant A, C_A	output1
Temperature of the reactor, T	output2
Process flow rate, q	100L/min
Inflow concentration of reactant A, C_{A0}	1mol/L
Inflow temperature, T_0	350.0K
Inlet coolant temperature, T_{c0}	350.0K
Reaction rate constant, k_0	$7.2 \times 10^{10} min^{-1}$
Specific heats, C_p, C_{pc}	1 cal/g/K
Density of reactant and the cooling fluid, ρ, ρ_c	$1 \times 10^3 g/L$
Reaction heat, $-\triangle H$	$-2 \times 10^5 cal/mol$
Heat transfer term, hA	$7 \times 10^5 cal/(min/K)$
The reactor volume, V	100L
Activation energy term, E/R	$1 \times 10^4 K$
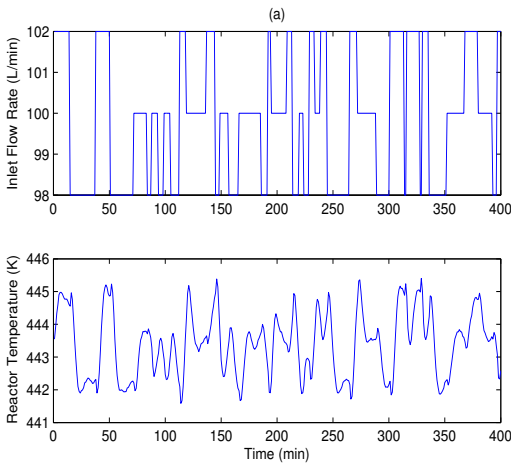
Fig. 4. Input-output data of the process, (a): Inlet flow rate (b): Reactor temperature

represent the state x_t for all t. The estimated parameter values after 50 iterations are given in Table IV.

It can be seen from the above estimation result that, all parameters converge to the neighborhood of the true values after certain iterations.

VI. CONCLUSION

This paper described a constrained Bayesian approach for identifying nonlinear state space model within the framework of the EM algorithm. Constrained particle filters approximation is performed in the expectation step. The capability of the proposed algorithm in handling missing observations is demonstrated through numerical examples. It has been shown that as the proportion of the missing part increases, more iterations are needed for the EM algorithm

TABLE IV

ESTIMATED PARAMETERS AFTER 50 ITERATIONS

Proportion of missing output	θ_1	$\theta_2 \times 10^{-2}$
10%	0.9703	1.9894
20%	0.9719	2.1002
30%	0.9709	1.9801

to achieve the convergence. In terms of computational efficiency, the proposed algorithm requires less CPU time than the comparative method.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] P. K. Pearson, Control Systems, Identification, *Encyclopedia of Physical Science and Technology*, 2004, pp 687-707.
[2] B. F. Feeny, NONLINEAR SYSTEM IDENTIFICATION, *Encyclopedia of Vibration*, 2004, pp 924-928.
[3] S. Khatibisepehr and B. Huang, Dealing with Irregular Data in Soft Sensors: Bayesian Method and Comparative Study, *Industrial & Engineering Chemistry Research*, vol. 22, 2008, pp 8713-8723.
[4] R. Wallin, A. J. Isaksson and L. Ljung, An iterative method for identification of ARX models from inceomplete data, *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, December 2000.
[5] A. P. Demspter, N. M. Larid and D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, vol. 39, 1977, pp 1-38.
[6] R. H. Shumway and D. S. Stoffer, An Approach to Time Series Smoothing and Forcasting Using the em Algorithm, *Journal of Time Series Analysis*, vol. 3, 1982, pp 253-264.
[7] A. Doucet, N. de Freitas and N. Gordon, Sequential Monte Carlo Methods in Practice, *Springer*, New York, 2001.
[8] R. B. Gopaluni, A Particle Filter Approach to Identification of Nonlinear Process under Missing Observations, *The Canadian Journal of Chemical Engineering*, vol. 86, 2008, pp 1081-1092.
[9] J. Kim, Ph.D Thesis, University of Pittsburgh, Pittsburgh, 2005.
[10] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, A Tutorial on Particle Filters for Onine Nonlinear/Non-Guassian Bayesian Tracking, *IEEE Transaction on Signal Processing*, vol. 50, 2002, pp 174-188.
[11] N. Bergman, Recursive Bayesian estimation: Navigation and tracking applications, *Ph.D Dissertation*, Linkping Univ., Linkping, Sweden, 1999.
[12] A. Doucet, On sequential Monte Carlo methods for Bayesian filtering, Dept. Eng., Univ. Cambridge, UK, Technical Report, 1998.
[13] J. S. Liu and R. Chen, Sequential Monte Carlo methods for dynamical systems, *Journal of the American Statistical Association*, vol. 93, 1998, pp 1032-1044.
[14] L. Lang, W. Chen, B. R. Bakshi, P. K. Goel and S. Ungarala, Bayesian estimation via sequential Monte Carlo sampling-Constrained dynamic systems, *Automatica*, vol. 43, 2007, pp 1615-1622.
[15] X. Shao, B. Huang and J. M. Lee, Constrained Bayesian state estimation - A comparative study and a new particle filter based approach, *Journal of Process Control*, vol. 20, 2010, pp 143-157.
[16] G. C. Goodwin and J. C. Aguero, Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models, *Proceedings of the 44th IEEE Conference on Decision and Control*, Seville, Spain, 2005.
[17] P. Vachhani, S. Narasimhan and R. Rengaswamy, Robust and reliable estimation via unscented recursive nonlinear dynamic data reconcillation, *Journal of Process Control*, vol. 16, 2006, pp 1075-1086.
[18] S. Ungarala, K. Li and Z. Chen, Constrained Bayesian state estimation using a cell filter, *Industrial & Engineering Chemistry Research*, vol. 47, 2008, pp 7312-7322.
[19] R. Senthil, K. Janarthanan, J. Prakash, Nonliear state estimation using fuzzy kalman filter, *Industrial & Engineering Chemistry Research*, vol. 25, 2006, pp 8678-8688.
[20] Z. Xu, J. Zhao, J. Qian and Y. Zhu, Nonliear MPC using identified LPV model, *Industrial & Engineering Chemistry Research*, vol. 6, 2009, pp 3043-3051.