Computers and Chemical Engineering 32 (2008) 3011–3022

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models

M.S. Varziri[a], A.A. Poyton[a], K.B. McAuley[a,*], P.J. McLellan[a], J.O. Ramsay[b]

[a] Department of Chemical Engineering, Queen's University, Kingston, ON, Canada K7L 3N6
[b] Department of Psychology, McGill University, Montreal, PQ, Canada H3A 1B1

ARTICLE INFO

Article history:
Received 11 December 2006
Received in revised form 31 March 2008
Accepted 14 April 2008
Available online 29 April 2008

Keywords:
Principal differential analysis
Parameter estimation
Dynamic models
Stochastic disturbances
B-splines

ABSTRACT

Iteratively refined principal differential analysis (iPDA) is a spline-based method for estimating parameters in ordinary differential equation (ODE) models. In this article we extend iPDA for use in differential equation models with stochastic disturbances and we demonstrate the probabilistic basis for the iPDA objective function using a maximum likelihood argument. This development naturally leads to a method for selecting the optimal weighting factor in the iPDA objective function. We demonstrate the effectiveness of iPDA using a simple two-output continuous-stirred-tank-reactor example, and we use Monte Carlo simulations to show that iPDA parameter estimates are superior to those obtained using traditional nonlinear least squares techniques, which do not account for stochastic disturbances.

1. Introduction

Parameter estimation in mathematical models is an important, difficult, and ubiquitous problem in chemical engineering and in many other areas of applied science. Fundamental process models can be exploited by many process optimization and control technologies (Biegler & Grossman, 2004), but it is important that appropriate parameter values are used so that model predictions match the underlying process behaviour. Obtaining good parameter values requires informative data for parameter estimation, as well as reliable parameter estimation techniques.

It is particularly difficult to estimate parameters in ordinary differential equation (ODE) models. The weighted sum of squared prediction errors is the usual minimization criterion for parameter estimation, and evaluating this criterion requires (numerical) solution of the ODEs. Sensitivity information, used by gradient-based parameter-estimation techniques, requires the solution of sensitivity equations (Leis & Kramer, 1988) or numerous additional simulations using perturbed parameter values. Numerical overflow and stability problems can arise when poor initial or intermediate parameter values are used during the course of parameter estimation (Biegler & Grossman, 2004).

A variety of ODE parameter estimation techniques have been used, ranging from traditional least-squares methods combined with repeated solution of differential equations (Bard, 1974; Bates & Watts, 1988; Ogunnaike & Ray, 1994; Seber & Wild, 1989) to multiple shooting (Bock, 1981, 1983), collocation-based methods (Biegler, 1984), and algorithms that use spline functions (Vajda & Valko, 1986; Varah, 1982). Biegler and Grossman (2004) provide a detailed survey of the existing methods.

In an attempt to develop an efficient and easy-to-use algorithm for estimating parameters in ODE models, Poyton, Varziri, McAuley, McLellan, and Ramsay (2006) proposed iteratively refined principal differential analysis (iPDA), which builds upon ideas from principal differential analysis (PDA). PDA is a functional data analysis tool that was proposed by Ramsay (1996) for empirical modelling using linear ODEs. PDA makes use of basis functions (usually B-splines) for estimating ODE parameters (Poyton et al., 2006; Ramsay, 1996; Ramsay & Silverman, 2005).

In this paper we will address some of the issues raised by Poyton et al. (2006) regarding iPDA. Most importantly we extend the use of iPDA to cases in which process noise (unmeasured disturbances that pass through the process) and measurement noise are both present, and we propose a criterion for selecting optimal weighting factors in the iPDA algorithm. We begin with a brief review of the iPDA algorithm and its advantages and shortcomings. Then we describe how iPDA can be used to estimate parameters in differential equation models with process disturbances (stochastic differential equation models (Maybeck, 1979)).

* Corresponding author. Tel.: +1 613 533 2768; fax: +1 613 533 6637.
E-mail address: kim.mcauley@chee.queensu.ca (K.B. McAuley).

0098-1354/$ – see front matter
doi:10.1016/j.compchemeng.2008.04.005

Finally, we use a simple continuous-stirred-tank-reactor example to compare parameter estimates obtained using iPDA with those obtained using a traditional nonlinear least squares (NLS) approach (Ogunnaike & Ray, 1994), which does not account for the stochastic process disturbances.

1.1. iPDA algorithm

To explain the iPDA algorithm, we will use the following simple first-order single-input single-output (SISO) nonlinear ODE model:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), \boldsymbol{\theta})\\ x(t_0) &= x_0\\ y(t_{mj}) &= x(t_{mj}) + \varepsilon(t_{mj})\end{aligned} \tag{1}$$

x is the state variable, u is the input variable and y is the output variable (which is the same as the state variable in this case). f is a nonlinear function of the vector of model parameters $\boldsymbol{\theta}$, state variables, and input variables. ε is a zero-mean uncorrelated random variable with variance σ_m^2.

The *first step* in iPDA is to fit a B-spline curve to the observed data. This empirical B-spline model is of the form:

$$x_{\sim}(t) = \sum_{i=1}^{c} \beta_i \phi_i \tag{2}$$

where $\beta_i, i = 1,\ldots, c$ are B-spline coefficients and $\phi_i(t)\ i = 1,\ldots, c$ are B-spline basis functions, for which a knot sequence must be specified. Please refer to Poyton et al. (2006) for a short introduction to B-splines and to de Boor (2001) for a detailed treatment. Note that Eq. (2) can be written in matrix form:

$$x_{\sim}(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\beta} \tag{3}$$

where $\boldsymbol{\varphi}(t)$ is a vector containing the c basis functions and $\boldsymbol{\beta}$ is vector of c spline coefficients. Note that the "$\sim$" subscript is used to imply an empirical curve that can be easily differentiated:

$$\dot{x}_{\sim}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{c} \beta_i \phi_i(t)\right) = \sum_{i=1}^{c} \beta_i \dot{\phi}_i(t) = \dot{\boldsymbol{\varphi}}^T \boldsymbol{\beta} \tag{4}$$

The empirical function $x_{\sim}(t)$ is determined by selecting the spline coefficients $\boldsymbol{\beta}$ that minimize the following objective function, given the most recent estimates for the fundamental model parameters $\boldsymbol{\theta}$:

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{j=1}^{n} (y(t_{mj}) - x_{\sim}(t_{mj}))^2 + \lambda \int_{t_0}^{t_{mn}} (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 \, \mathrm{d}t \right\} \tag{5}$$

where n is the number of data points and t_{mj} is the time at which the jth data point was measured. We will refer to the first term $\sum (y(t_{mj}) - x_{\sim}(t_{mj}))^2$ in the objective function as SSE (the sum of squared prediction errors) and to the second term $\int (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 \, \mathrm{d}t$ as PEN (the model-based penalty). PEN is a measure of how well the empirical curve satisfies the ODE model. In the initial iteration of iPDA, initial guesses for the fundamental parameters $\boldsymbol{\theta}$ are required to compute PEN. Optimal B-spline coefficients $\boldsymbol{\beta}$ are obtained by considering the measured observations and also the extent to which the empirical curve satisfies the model. The model-based penalty prevents B-spline curves that make the SSE small, but are inconsistent with the behaviour of the fundamental model. The positive weighting factor λ determines how the empirical B-spline curve balances between matching the observed data and satisfying the ODE model. A small λ is appropriate when we believe the measurements more than the model and a large λ is appropriate when we have confidence in our model but our measured observations are noisy. Poyton et al. (2006) pointed out that proper selection of λ is of crucial importance and showed that the quality of the B-spline curves (and also the parameter estimates $\hat{\boldsymbol{\theta}}$ obtained using iPDA) depends very much on the value of λ. In this paper we propose a means of determining an optimal λ given some knowledge about measurement noise and model disturbances.

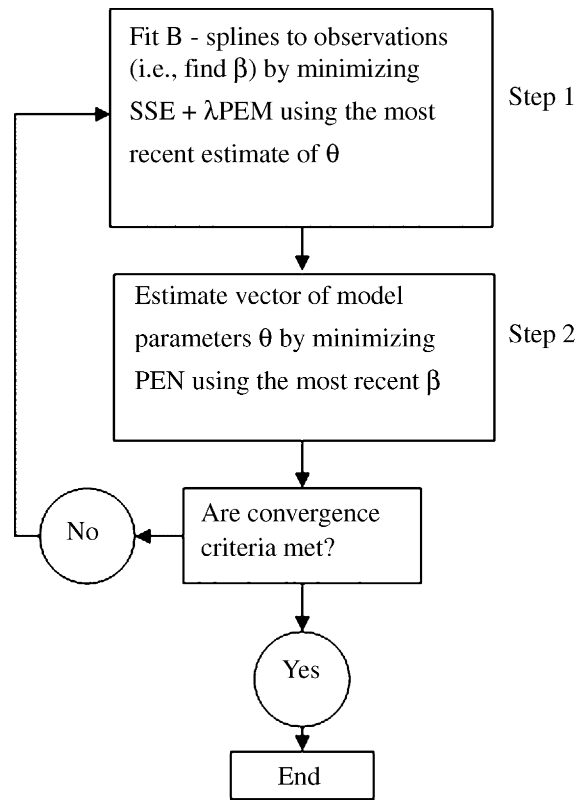

Fig. 1. iPDA algorithm.

The *second step* in iPDA is to estimate the vector of fundamental model parameters $\hat{\boldsymbol{\theta}}$, using fixed values of the B-spline coefficients $\hat{\boldsymbol{\beta}}$ (and hence fixed $x_{\sim}$) from step one. The fundamental model parameters are selected to minimize the following objective function:

$$\min_{\boldsymbol{\theta}} \int_{t_0}^{t_{mn}} (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 \, \mathrm{d}t \tag{6}$$

Next, we return to the first step and re-estimate the B-spline coefficients using $\hat{\boldsymbol{\theta}}$ obtained from step two. iPDA iterates between step one and step two until convergence, as shown in Fig. 1.

The original PDA algorithm (Ramsay, 1996) is not iterative, and it uses penalties on higher-order derivatives (e.g., $\int (\ddot{x}(t))^2 \, \mathrm{d}t$) to prevent over-fitting of the data, instead of the model-based penalty of iPDA. PDA has been used in various areas such as handwriting analysis (Ramsay, 2000), analysis of movement of the lips during speech (Lucero, 2002; Ramsay & Munhall, 1996), economic modelling (Ramsay & Ramsey, 2002) and meteorological modelling (Ramsay & Silverman, 2005). Poyton et al. (2006) showed that poor spline fits from the first step of the original PDA algorithm give misleading derivative information, which results in inaccurate parameter estimates from the second step. They showed that inclusion of the model-based penalty in the first step of iPDA, along with its iterative nature, eliminates this problem.

Since $\sum (y(t_{mj}) - x_{\sim}(t_{mj}))^2$ is constant when $\hat{\boldsymbol{\beta}}$ is fixed, step two of iPDA is equivalent to minimizing:

$$\min_{\boldsymbol{\theta}} \left\{ \sum_{j=1}^{n} (y(t_{mj}) - x_{\sim}(t_{mj}))^2 + \lambda \int (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 \, dt \right\} \tag{7}$$

As a result, when the estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ have converged, the following overall objective function is minimized:

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}} \left\{ \sum_{j=1}^{n} (y(t_{mj}) - x_{\sim}(t_{mj}))^2 + \lambda \int (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 \, dt \right\} \tag{8}$$

This optimization problem can be solved simultaneously for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, instead of using the iterative two-step procedure described above. Since the vector of spline coefficients, is generally of high dimension, Eq. (8) is the objective function for a large-scale nonlinear minimization problem. One benefit of the crude iterative approach shown in Fig. 1 is that it can simplify this large-scale nonlinear problem: If the ODE is linear in the inputs and outputs, then the first step of iPDA is a large *linear* least-squares problem, and the second step of iPDA, which is a nonlinear least-squares problem is of much smaller dimension. Nonetheless, we believe that alternative approaches for solving large-scale minimization methods (Biegler, 1984) should be investigated for obtaining the iPDA parameter estimates.

1.2. iPDA advantages, shortcomings and the purpose of current paper

Since the empirical B-spline curve that is fitted to the observations can be easily differentiated with respect to time, iPDA circumvents the need for repeated numerical solution of ODEs, which is required by traditional NLS methods (Ogunnaike & Ray, 1994). Solving ODEs numerically during NLS estimation can sometimes lead to numerical overflow and instabilities, especially when the initial guesses (or along-the-way estimates) of model parameters are poor, or if the dynamic model contains unstable modes (Ascher, Mathheij, & Russell, 1988; Bard, 1974; Li, Osborne, & Pravan, 2005; Tanartkit & Biegler, 1995). These problems are not encountered by iPDA.

Another advantage of iPDA arises from the form of the objective function for the parameter estimation step (step two). Since the integral of the squared deviation for the *differentiated* form of the model is minimized in Eq. (6), as opposed to the sum of squared prediction errors as in traditional NLS, sensitivity information is readily available. Analytical derivatives can be used because $\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta})$ can easily be differentiated with respect to $\boldsymbol{\theta}$. So, unlike traditional NLS, there is no need to numerically integrate sensitivity equations (Bard, 1974; Bates & Watts, 1988; Seber & Wild, 1989). Objective function (6) has a further advantage in that the nonlinearity of parameters is often less severe in the differentiated form of the model than in the integrated solution (e.g., kinetic rate constants and heat-transfer coefficients often appear linearly on the right-hand-side of the ODE, but would appear in exponential terms in the integrated response). Therefore, iPDA may be less susceptible to problems associated with poor initial parameter guesses than are traditional methods.

iPDA is particularly well suited to parameter estimation in ODE models in which some or all of the initial conditions for the states are unknown (boundary value problems). When using iPDA, there is no need to repeatedly solve the ODEs with different guesses for the initial conditions. Estimates for initial conditions are provided automatically by $x_{\sim}(0)$. As we will show using the examples in this article, it is straightforward to incorporate known initial conditions in the B-spline curve-fitting step by fixing the value of one of the spline coefficients. Parameter estimation problems involving state-variable constraints (Biegler & Grossman, 2004) could also be readily handled using iPDA.

Using basis functions, either to empirically fit the observations or to approximate the solution of the ODEs (collocation-based methods), during parameter estimation, is not exclusive to iPDA (e.g., Benson, 1979; Biegler, 1984; Logsdon & Biegler, 1992; Swartz & Bremermann, 1975; Tang, 1971; Vajda & Valko, 1986; Varah, 1982). Several types of basis functions have been used for discretizing ODE models during parameter estimation (Biegler & Grossman, 2004). For example, Biegler (1984) used Lagrange interpolating polynomials because they facilitate providing bounds and starting points for coefficients that are to be estimated. B-spline functions were selected for our iPDA algorithm (and for original PDA) because they are bounded polynomials that are non-zero only over a finite interval. B-splines provide "compact support" for the empirical curve (de Boor, 2001), which leads to banded matrices that are numerically attractive for smoothing and inverse problems (Eilers & Marx, 1996; O'Sullivan, 1986; Ramsay & Silverman, 2005).

A further benefit of iPDA over other basis-function methods is that the model-based penalty (PEN) in the B-spline fitting objective function (Eq. (5)) regularizes the fitted curve and prevents it from having unrealistic features that are not consistent with the model. Because of this property, iPDA is an algorithm in the class of regularization methods, which are widely used for solving linear and nonlinear inverse problems (Binder, Blank, Dahmen, & Marquardt, 2002; Kirsch, 1996; O'Sullivan, 1986). Note that the model-based penalty (PEN) in iPDA is not a hard constraint, but rather a soft constraint that is only satisfied to some extent, which is determined by the value of the weighting factor λ. In other collocation-based methods the parameter estimation problem is posed as a hard-constrained minimization problem, where the sum of squared prediction errors (SSE) is minimized subject to the discretized ODE (Biegler & Grossman, 2004). As suggested by Poyton et al. (2006), imposing the discretized ODE as a soft constraint in iPDA may be particularly advantageous for estimating parameters in models in which unmeasured stochastic disturbances influence dynamic process behaviour. In this paper we consider these types of models and we demonstrate how iPDA readily addresses the resulting parameter estimation problem.

Another issue raised by Poyton et al. (2006) is whether or not iPDA can be used for problems in which some of the states are not measured. Our recent investigations confirm that the answer is yes; iPDA can readily handle estimation problems with unmeasured states so long as certain observability conditions are met (Varziri, 2008), which are analogous to the conditions required for estimation of unmeasured states using an extended Kalman filter.

Shortcomings of iPDA as listed by Poyton et al. (2006) are as follows:

- iPDA requires an appropriate B-spline knot sequence. The quality of the parameter estimates $\hat{\boldsymbol{\theta}}$ depends on the empirical B-spline curve which, in turn, relies on the selected knot sequence. Optimal knot placement is currently under investigation in our research work. However, it seems that if enough data points are available, placing one knot at each observation point (as will be shown in the case study) can lead to satisfactory results. Additional knots maybe required when there are sharp changes in the output. Using too many knots can lead to long computational times.
- iPDA parameter estimates depend on the weighting factor λ in Eq. (5). Heuristically, the weighting factor should depend on

measurement uncertainties and model disturbances. The more uncertain the measurements are, the larger λ should be and the more uncertain the model is, the smaller λ should be. The uncertainty in the model (due to unmodeled disturbances or other model imperfections) can have several sources. One possible source of model disturbances is uncertainty in the inputs to the system. Traditional NLS assumes that the inputs to the system are perfectly known, however in practice, due to flaws in measurement devices and valves and also because of external unmeasured or unmodeled disturbances, this is rarely the case. Finally, model uncertainty arises because there are some physical phenomena that have not been included in the model; in other words there may be missing or improperly specified terms in the model. In the current article we add a stochastic term to the right-hand-side of the ODE in Eq. (1) to account for model uncertainties. We then show that minimizing the objective function in Eq. (5) leads to optimal B-spline coefficients for smoothing the observations, taking into account the levels of process noise and measurement noise in the system. This development naturally leads to an expression for the optimal value of the weighting factor λ.

- No confidence interval expressions have been developed to assist modelers in making inferences about the quality of parameter estimates and model predictions obtained using iPDA. We will address this issue in our ongoing iPDA research. Also, iPDA has not yet been used for parameter estimation in larger-scale parameter estimation problems. Further research will be required to determine efficient algorithms for obtaining iPDA parameter estimates.

2. iPDA in presence of model disturbance

To keep the notation compact we use a single-input single-output (SISO) nonlinear model; the multi-input multi-output (MIMO) case is presented in Appendix A.

Consider the following continuous-time stochastic dynamic model (Astrom, 1970; Brown & Hwang, 1992; Maybeck, 1979):

$$\dot{x}(t) = f(x(t), u(t), \boldsymbol{\theta}) + \eta(t)$$
$$x(t_0) = x_0 \qquad (9)$$
$$y(t_{mj}) = x(t_{mj}) + \varepsilon(t_{mj})$$

The initial condition, x_0 is a normally distributed random variable with mean $E\{x_0\}$ and variance σ_0^2. $\eta(t)$ is a continuous zero-mean stationary white-noise process with covariance matrix $E\{\eta(t)\eta(t+\tau)\} = Q\delta(\tau)$, where Q is the corresponding power spectral density and $\delta(.)$ is the Dirac delta function. For the discrete-time white-noise process (Maybeck, 1979):

$$E\{\eta(j_1\ \Delta t)\eta(j_2\,\Delta t)\} = \begin{cases} \dfrac{Q}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \qquad (10)$$

where j_1 and j_2 are integers and Δt is the sampling period. A discrete random white noise sequence as shown in Fig. 2 is a series of step functions with sampling interval Δt, where the variance of the white noise, $\sigma_p^2 = Q/\Delta t$. In the limiting case where $\Delta t \to 0$ we get the same behaviour as in the continuous case (using the Dirac delta function). All the other terms in (9) remain the same as those in (1). We also assume that the process noise $\eta(t)$ and the measurement noise $\varepsilon(t)$ are not correlated. In the next section we use Bayesian arguments to justify the use of the B-spline-fitting objective function in (5) and we prescribe a method for selecting an optimal λ given Q and σ_m^2.

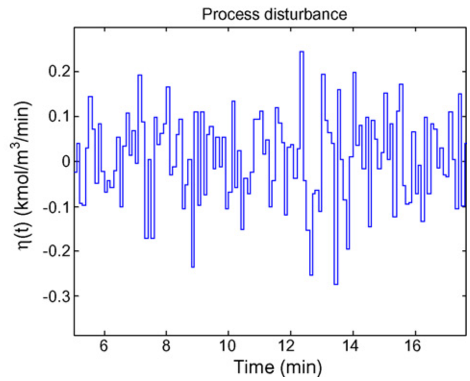

Fig. 2. Process disturbance.

2.1. Selecting the optimal weighting factor

We will show that

$$\lambda_{\text{opt}} = \frac{\sigma_m^2}{Q} \qquad (11)$$

where the subscript "opt" indicates the optimal weighting factor, which will lead to maximum likelihood parameter estimates for $\boldsymbol{\beta}$, given $\boldsymbol{\theta}$. The resulting optimal spline curves $x_{\sim}$ will lead to optimal values of $\hat{\boldsymbol{\theta}}$ (the estimate of $\boldsymbol{\theta}$) after the algorithm converges.

Before we begin the mathematical derivations, it is helpful to outline the approach that we will use. The state variable x in (9) is a random variable (due to the stochastic input $\eta(t)$), which evolves in time. Values of x, sampled at various times, have a prior joint distribution, which is multivariate normal due to the assumptions about $\eta(t)$. Once the measured data become available, the posterior joint distribution of the sampled values of x and the measured response, y, can be obtained. We will consider the general case when the initial value x_0 is not perfectly known, and then we will demonstrate that maximizing the likelihood of the joint distribution of the sampled state values and measured observations given the parameter values, is equivalent to minimizing the iPDA objective function (5) in the restricted case when we assume that x_0 in (9) is perfectly known.

At the discrete time $t_i = t_{i-1} + \Delta t$, where the sampling interval Δt is small, Eq. (9) can be written using the following Euler approximation:

$$x(t_{i-1} + \Delta t) = x(t_i) = x(t_{i-1}) + f(x(t_{i-1}), u(t_{i-1}), \boldsymbol{\theta})\Delta t + \eta(t_{i-1})\Delta t \qquad (12)$$

Consider $x(t_i)$ at $q+1$ uniformly spaced time points, t_i, $i = 0,\ldots, q$ so that $q\Delta t = T$, where $T = t_q - t_0$ is the overall time span for the model predictions. For brevity, we define $x_i = x(t_i)$. Please note that the set of times at which the measurements are available is a subset of t_i $(i = 0,\ldots, q)$ and is denoted by t_{mj} $(j = 1,\ldots, n)$. The measurement times t_{mj} do not need to be uniformly spaced. The vector of outputs at observation times $y(t_{mj})$ $(j = 1,\ldots, n)$ and its corresponding state vector of true values $x(t_{mj})$ $(j = 1,\ldots, n)$ and measurement noise vector $\varepsilon(t_{mj})$ $(j = 1,\ldots, n)$ are denoted by $\mathbf{y}_m$, $\mathbf{x}_m$, and $\boldsymbol{\varepsilon}_m$, respectively.

From Bayes' rule, the joint probability density of $x_0,\ldots, x_q$ given the vector of observations $\mathbf{y}_m$ can be written as

$$p(x_0, \ldots, x_q, \mathbf{y}_m|\boldsymbol{\theta}) = p(\mathbf{y}_m|x_0, \ldots, x_q, \boldsymbol{\theta}) \times p(x_0, \ldots, x_q|\boldsymbol{\theta}) \qquad (13)$$

Due to the Markov property (Gong, Wahba, Johnson, & Tribbia, 1998; Maybeck, 1979):

$$p(x_0, \ldots, x_q|\boldsymbol{\theta}) = p(x_q|x_{q-1}, \boldsymbol{\theta}) \times \cdots \times p(x_1|x_0, \boldsymbol{\theta}) \times p(x_0|\boldsymbol{\theta}) \qquad (14)$$

Substituting (14) into (13) gives:

$$p(x_0,\ldots,x_q,\mathbf{y}_m|\boldsymbol{\theta}) = p(\mathbf{y}_m|x_0,\ldots,x_q,\boldsymbol{\theta})\times p(x_q|x_{q-1},\boldsymbol{\theta})\times\cdots \times p(x_1|x_0,\boldsymbol{\theta})\times p(x_0|\boldsymbol{\theta}) \tag{15}$$

We now evaluate each term on the right-hand-side of (15). From: $y(t_{mj})=x(t_{mj})+\varepsilon(t_{mj})$

$$\mathbf{y}_m = \mathbf{x}_m + \boldsymbol{\varepsilon}_m \tag{16}$$

where $\mathbf{x}_m$ is the vector of true state values at the measurement times. Therefore, $p(\mathbf{y}_m|x_0,\ldots,x_q,\boldsymbol{\theta})$, is a multivariate Gaussian distribution:

$$p(\mathbf{y}_m|x_0,\ldots,x_q,\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}\sigma_m^n}\exp\left\{\frac{-(\mathbf{y}_m-\mathbf{x}_m)^T(\mathbf{y}_m-\mathbf{x}_m)}{2\sigma_m^2}\right\} \tag{17}$$

with mean $E\{\mathbf{y}_m|x_0,\ldots, x_q, \boldsymbol{\theta}\}=\mathbf{x}_m$ and covariance matrix, $\mathrm{cov}\left\{\mathbf{y}_m|x_0,\ldots,x_q\right\}=\sigma_m^2\mathbf{I}_{n\times n}$.

From, Eq. (12), $p(x_i|x_{i-1})$ is a Gaussian distribution:

$$p(x_i|x_{i-1},\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sqrt{Q\,\Delta t}}\exp\left\{\frac{-(x_i-E\{x_i|x_{i-1},\boldsymbol{\theta}\})^2}{2Q\,\Delta t}\right\} \tag{18}$$

with

$$E\{x_i|x_{i-1},\boldsymbol{\theta}\} = x_{i-1}+f(x_{i-1},u_{i-1},\boldsymbol{\theta})\Delta t \quad \text{and}$$

$$\mathrm{cov}\{x_i|x_{i-1},\boldsymbol{\theta}\} = \mathrm{cov}\{\eta_{i-1}\,\Delta t\} = \left(\frac{Q}{\Delta t}\right)\Delta t^2 = Q\,\Delta t \tag{19}$$

We assume that the initial condition x_0 has a Gaussian distribution:

$$p(x_0|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{\frac{-(x_0-E\{x_0|\boldsymbol{\theta}\})^2}{2\sigma_0^2}\right\} \tag{20}$$

Therefore, from Eqs. (15), (17), (18), and (20):

$$p(x_0,\ldots,x_q,\mathbf{y}_m|\boldsymbol{\theta}) = \exp\left\{\frac{-(\mathbf{y}_m-\mathbf{x}_m)^T(\mathbf{y}_m-\mathbf{x}_m)}{2\sigma_m^2}\right\} \times\prod_{i=1}^{q}\exp\left\{\frac{-(x_i-E\{x_i|x_{i-1},\boldsymbol{\theta}\})^2}{2Q\Delta t}\right\} \times\exp\left\{\frac{-(x_0-E\{x_0|\boldsymbol{\theta}\})^2}{2\sigma_0^2}\right\} \tag{21}$$

The optimal state and parameter estimates are denoted $\hat{\mathbf{x}}=\hat{x}_0,\ldots,\hat{x}_q$ and $\hat{\boldsymbol{\theta}}$, respectively and minimize:

$$\frac{(\mathbf{y}_m-\mathbf{x}_m)^T(\mathbf{y}_m-\mathbf{x}_m)}{\sigma_m^2} + \sum_{i=1}^{q}\frac{1}{Q}\left(\frac{x_i-E\{x_i|x_{i-1},\boldsymbol{\theta}\}}{\Delta t}\right)^2\Delta t + \frac{(x_0-E\{x_0|\boldsymbol{\theta}\})^2}{\sigma_0^2} \tag{22}$$

which is the natural logarithm of the right-hand-side of Eq. (21) multiplied by −1.

In the limiting case where $\Delta t\to 0$ (Jazwinski, 1970), Eq. (19) implies

$$\frac{x_i-E\{x_i|x_{i-1},\boldsymbol{\theta}\}}{\Delta t} \to \dot{x}_{i-1}-f(x_{i-1},u_{i-1},\boldsymbol{\theta}) \tag{23}$$

and assuming that $E\{x_0|\boldsymbol{\theta}\}=E\{x_0\}$ Eq. (22) becomes

$$\frac{(\mathbf{y}_m-\mathbf{x}_m)^T(\mathbf{y}_m-\mathbf{x}_m)}{\sigma_m^2} + \frac{1}{Q}\int_{t_0}^{t_q}(\dot{x}(t)-f(x(t),u(t),\boldsymbol{\theta}))^2\,\mathrm{d}t + \frac{(x_0-E\{x_0\})^2}{\sigma_0^2} \tag{24}$$

If the initial condition x_0 is perfectly known, the last term in Eq. (24) vanishes. If we assume that $x(t)$ can be approximated by B-spline curves so that $x(t)\cong x_\sim(t)=\boldsymbol{\varphi}^T(t)\boldsymbol{\beta}$, then minimizing (24) is equivalent to minimizing:

$$\sum_{j=1}^{n}(y(t_{mj})-x_\sim(t_{mj}))^2 + \frac{\sigma_m^2}{Q}\int_{t_0}^{t_q}(\dot{x}_\sim(t)-f(x_\sim(t),u(t),\boldsymbol{\theta}))^2\,\mathrm{d}t \tag{25}$$

Eq. (25) is the same as the iPDA objective function for B-spline fitting, Eq. (5), with $\lambda=\sigma_m^2/Q$. As a result, we have shown that optimal B-spline coefficients, which result in $x_\sim$ approximating the true curve for x, are obtained using the iPDA weighting coefficient $\lambda_{\mathrm{opt}}=\sigma_m^2/Q$.

If the integral in (25) is approximated by discrete sums:

$$\sum_{j=1}^{n}(y(t_{mj})-x_\sim(t_{mj}))^2 + \frac{\sigma_m^2}{\sigma_p^2}\sum_{i=1}^{q}(\dot{x}_\sim(t_i)-f(x_\sim(t_i),u(t_i),\boldsymbol{\theta}))^2 + \frac{\sigma_m^2}{\sigma_0^2}(x_0-E\{x_0\})^2 \tag{26}$$

where $\sigma_p^2=Q/\Delta t$ is the process noise variance, so that the optimal weighting factor using a discretized model and discrete process noise is

$$\lambda_{\mathrm{opt-discrete}} = \frac{\sigma_m^2}{\sigma_p^2} \tag{27}$$

If the modeler has knowledge about σ_m^2 from replicate measurements of the outputs, and about σ_p^2 from uncertainties in input settings, and the types and magnitudes of anticipated disturbances, then a reasonable value of λ can be selected. The information required for optimal selection of λ is analogous to the information required to tune a Kalman filter (Gagnon & MacGregor, 1991).

Traditional NLS parameter estimation corresponds to optimizing the iPDA objective function in (8) in the limiting case when $\lambda\to\infty$, because traditional NLS methods assume that $\sigma_p^2=0$. In this case, the B-spline curve tends to satisfy the differential equation model perfectly (assuming there are sufficient spline knots) and the fundamental parameter estimates, and the SSE is minimized through optimal selection of $\hat{\boldsymbol{\theta}}$ (which influences $x_\sim$ as the iterations proceed.

In Appendix A, we extend these results to the general multivariate case. To illustrate what happens to the iPDA objective function in a simple multivariate estimation problem, we consider the following nonlinear dynamic system with two measured outputs:

$$\begin{cases} \dot{x}_1(t)=f_1(x_1(t),x_2(t),u_1(t),u_2(t),\boldsymbol{\theta})+\eta_1(t)\\ \dot{x}_2(t)=f_2(x_1(t),x_2(t),u_1(t),u_2(t),\boldsymbol{\theta})+\eta_2(t)\\ y_1(t_{m1j})=x_1(t_{m1j})+\varepsilon_1(t_{m1j})\\ y_2(t_{m2j})=x_2(t_{m2j})+\varepsilon_2(t_{m2j}) \end{cases} \tag{28}$$

For this system, the matrices defined in Eq. (36) of Appendix A are

$$\mathbf{x}(t)=\begin{bmatrix}x_1(t)\\x_2(t)\end{bmatrix},\ \mathbf{F}=\begin{bmatrix}f_1\\f_2\end{bmatrix},\ \mathbf{u}(t)=\begin{bmatrix}u_1(t)\\u_2(t)\end{bmatrix},\ \mathbf{C}=\begin{bmatrix}1&0\\0&1\end{bmatrix},$$

$$\mathbf{E}=\begin{bmatrix}\sigma_{m1}^2&0\\0&\sigma_{m2}^2\end{bmatrix},\ \text{and}\ \mathbf{Q}=\begin{bmatrix}Q_{p1}&0\\0&Q_{p2}\end{bmatrix}.$$

The resulting iPDA objective function for optimal spline fitting (from Eq. (47) in Appendix A) is:

$$\frac{1}{\sigma_{m1}^2}\sum_{j=1}^{N_1}(y_1(t_{m1j})-x_{\sim 1}(t_{m1j}))^2+\frac{1}{\sigma_{m2}^2}\sum_{j=1}^{N_2}(y_2(t_{m2j})-x_{\sim 2}(t_{m2j}))^2$$
$$+\frac{1}{Q_{p1}}\int(\dot{x}_{\sim 1}(t)-f_1(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta}))^2\,dt$$
$$+\frac{1}{Q_{p2}}\int(\dot{x}_{\sim 2}(t)-f_2(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta}))^2\,dt \quad (29)$$

where N_1 and N_2 are the number of available measurements for outputs y_1 and y_2, respectively. This multivariate iPDA objective function readily accommodates outputs that are measured using different sampling rates and measurements that are made using irregularly spaced sampling times. The two SSE terms that appear in this objective function are each weighted by the reciprocal of the variance for their respective measurements. The objective function also contains two PEN terms corresponding to the integral of the squared deviations in the two stochastic differential equations. Because the development in Appendix A assumes that the noise variables $\eta_1(t)$ and $\eta_2(t)$ are independent, no cross-product term (involving both $\dot{x}_1$ and $\dot{x}_2$) appears in the objective function. Each of the PEN terms is weighted by reciprocal of its respective process noise variance, which becomes more readily apparent if the integrals are approximated by sums:

$$\frac{1}{\sigma_{m1}^2}\sum_{i=1}^{N_1}(y_1(t_{mi})-x_{\sim 1}(t_{mi}))^2+\frac{1}{\sigma_{m2}^2}\sum_{i=1}^{N_2}(y_2(t_{mi})-x_{\sim 2}(t_{mi}))^2$$
$$+\frac{1}{\sigma_{p1}^2}\sum_{i=0}^{q}(\dot{x}_{\sim 1}(t_i)-f_1(x_{\sim 1}(t_i),x_{\sim 2}(t_i),u_1(t_i),u_2(t_i),\boldsymbol{\theta}))^2$$
$$+\frac{1}{\sigma_{p2}^2}\sum_{i=0}^{q}(\dot{x}_{\sim 2}(t_i)-f_2(x_{\sim 1}(t_i),x_{\sim 2}(t_i),u_1(t_i),u_2(t_i),\boldsymbol{\theta}))^2 \quad (30)$$

where

$$\sigma_{p1}^2=\frac{Q_{p1}}{\Delta t}\text{ and }\sigma_{p2}^2=\frac{Q_{p2}}{\Delta t}.$$

Since each SSE term in the objective function is inversely proportional to its corresponding measurement error variance and each PEN term is inversely proportional to its corresponding model disturbance variance, it is straightforward to write an appropriate iPDA objective function for any multivariate problem. The main difficulty lies in obtaining estimates for these variances (or their ratios) in cases where all that is available is a dynamic model and some data. We are hopeful that the literature on tuning of Kalman filters (Maybeck, 1979) will provide some insight into this problem.

Objective functions in (29) and (30) reveal how iPDA can be conducted when some of the states are unmeasured. For example, if no measurements are available for output y_1 so that $N_1 = 0$, the first term in the objective functions disappears, and the spline curves $x_{1\sim}$ and $x_{1\sim}$ are fitted simultaneously using the remaining SSE term and the two PEN terms (Varziri, 2008). Then, step 2 of the iPDA algorithm involves selecting $\boldsymbol{\theta}$ to minimize the sum of the two PEN terms (Varziri, 2008).

3. Case study

In this case study we test our results for optimal selection of weighting factors in the iPDA objective function using two examples: a linear SISO continuous stirred tank reactor (CSTR), and a nonlinear MIMO CSTR. Note that we have also demonstrated the applicability of iPDA to more complicated problems including a nonlinear CSTR in which unmeasured states and nonstationary stochastic disturbances are present (Varziri, McAuley, & McLellan, 2008), and a nylon polymerization reactor model (Varziri, McAuley, & McLellan, submitted for publication-b). The algorithm has also been extended so that it can be applied to cases in which the process disturbance intensity, Q, is unknown (Varziri, McAuley, & McLellan, submitted for publication-a; Varziri et al., submitted for publication-b).

3.1. Linear SISO CSTR

First, we use a simple linearized continuous CSTR example from Poyton et al. (2006), with a stochastic term $\eta(t)$ added to the original model:

$$\frac{dC'_A(t)}{dt}=w_C C'_A(t)+w_T T'(t)+\eta(t)$$
$$C'_A(0)=0 \quad (31)$$
$$y(t_{mj})=C'_A(t_{mj})+\varepsilon(t_{mj})$$

where $C'_A = C_A - C_{As}$ and $T = T - T_s$ are the concentration of reactant A and the temperature, respectively, in deviation variables. The discrete process disturbance $\eta(t)$ used in our simulations (see Fig. 2) is a series of random step inputs whose duration (Δt = 0.1 min) is very short compared to the simulation time (24 min) and the process time constant. For this disturbance sequence in Fig. 2, $\sigma_p^2 = 2\times10^{-3}(\text{kmol/m}^3/\text{min})^2$. The measurement noise $\varepsilon(t_{mj})\,j = 1,\ldots, n$ is a white noise sequence with variance $\sigma_m^2 = 4\times10^{-4}(\text{kmol/m}^3)^2$.

The stochastic differential equation in Eq. (31) was obtained by linearizing the following nonlinear stochastic differential equation:

$$\frac{dC_A}{dt}=\frac{F}{V}(C_{A0}-C_A)-k_{\text{ref}}\exp\left(\frac{-E}{R}\left(\frac{1}{T}-\frac{1}{T_{\text{ref}}}\right)\right)C_A+\eta(t) \quad (32)$$

so the coefficients w_C and w_T in (31) depend on the unknown parameters k_{ref} (a kinetic rate constant) and E/R (an activation energy parameter), as well as the steady-state operating conditions:

$$w_C=-\frac{F_s}{V}-k_{\text{ref}}\exp\left(-\frac{E}{R}\left(\frac{1}{T_s}-\frac{1}{T_{\text{ref}}}\right)\right)$$
$$w_T=-k_{\text{ref}}\frac{E}{R}\frac{C_{As}}{T_s^2}\exp\left(-\frac{E}{R}\left(\frac{1}{T_s}-\frac{1}{T_{\text{ref}}}\right)\right).$$

We assume that $k_{\text{ref}} = 0.461\ \text{min}^{-1}$ and $E/R = 8330.1$ K are the true values of the parameters, that the feed rate is steady at $F_s = 0.05\ \text{m}^3\ \text{min}^{-1}$, the inlet reactant concentration C_{A0} is constant at 2.0 kmol m^{-3}, the reactor volume is $V = 1.0\ \text{m}^3$ and the reference temperature is $T_{\text{ref}} = 350$ K. When the reactor is operated at a constant temperature, $T_s = 332$ K, the resulting (expected) steady-state concentration of reactant A is $C_{As} = 0.567\ \text{kmol m}^{-3}$. We assume that the initial concentration in the reactor is C_{As} and that this initial concentration is known to the modeler. We also assume that the temperature control system is very effective, so that step changes in the temperature set point result in instantaneous step changes in the reactor temperature.

The parameters k_{ref} and E/R were estimated using the step change in temperature shown in Fig. 3, which produces the concentration response shown in Fig. 4. The true response of the concentration (obtained using the true parameter values and the true stochastic disturbance sequence $\eta(t)$) is shown as the dashed line in Fig. 4. The noisy measurements (241 equally spaced concentration measurements, with 10 measurements per minute) and the spline fit $C_{A\sim}(t)$ are also shown.

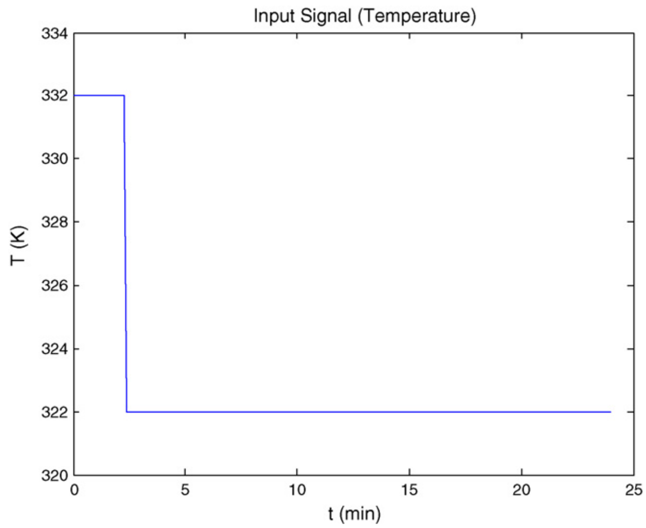

Fig. 3. Input signal.

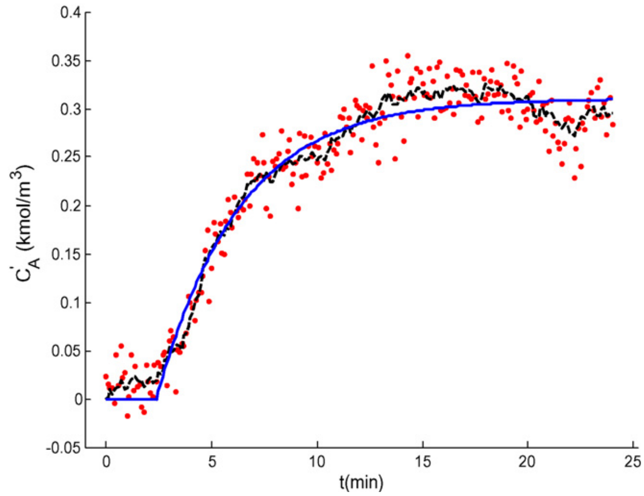

Fig. 5. Observed, true, and predicted response for NLS, linear SISO CSTR (●, simulated data; - - -, response of the system with true parameters and true stochastic noise; —, NLS response).

The objective function used to fit $C_{A\sim}(t)$ was:

$$\sum_{j=1}^{241}(C_A(t_{mj}) - C_{A\sim}(t_{mj}))^2 + \lambda \int_{t=0}^{t=24}\left(\frac{dC'_{A\sim}(t)}{dt} - w_C C'_{A\sim}(t) - w_T T'(t)\right)^2 dt \qquad (33)$$

where $C_{A\sim}(t)$ is the B-spline fit and the optimal value of λ is $\lambda_{opt} = \sigma_m^2/Q = \sigma_m^2/\sigma_p^2\,\Delta t = 2.0$ min (from Eq. (11)). In our objective function evaluations, we approximated the integral in Eq. (33) by discrete sums with Δt = 0.1 min, so that $\lambda_{opt-discrete} = \sigma_m^2/\sigma_p^2 = 0.2\,\text{min}^2$. In our simulations of the true response, we solved the linearized stochastic differential equation model using a Runge-Kutta method (ODE 45 in Simulink with relative tolerance 1E−3 and automatic absolute tolerance). In our subsequent example, the true plant is a more realistic nonlinear CSTR.

We obtained optimal parameter values in two different ways: (i) using the iterative procedure shown in Fig. 1 iPDA and (ii) by simultaneously estimating the combined vector of two fundamental model parameters and B-spline coefficients $\boldsymbol{\tau} = [\boldsymbol{\theta}', \boldsymbol{\beta}']'$ in Eq. (33) using "lsqnonlin" routine in Matlab. The parameter estimates from the simultaneous approach were on average better than the iterated method and hence preferred. When fitting the B-splines we used one knot at each observation time. We found that due to the stochastic disturbance, using coincident knots at the time when the step change in T occurred did not improve the overall fit to the observed data. Another reason is that placing one knot at each observation point in this example provides B-spline curves that are flexible enough to fit the observed data. The effect of coincident knots is more obvious when fewer knots are used.

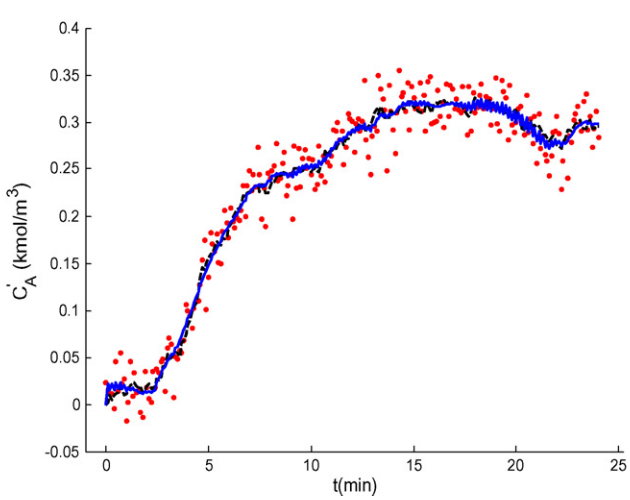

Fig. 4. Measured, true, and fitted responses for the linear SISO CSTR model obtained using iPDA with $\lambda_{opt-discrete} = 0.2\,\text{min}^2$. (●, simulated data; - - -, response of the system with true parameters and true stochastic noise; —, iPDA response).

Using the simulated data shown in Fig. 4, model parameters were also estimated using traditional nonlinear least squares estimation (ODE 45 in Simulink was used to repeatedly solve the ODE model with $\eta(t) = 0$ and the associated sensitivity equations). The fitted response is shown along with the data and the true process behaviour in Fig. 5. Note that the fitted model response obtained from traditional NLS is much further from the true response curve than is the iPDA state response curve $C_{A\sim}(t)$, which is shown in Fig. 4. When iPDA is used for parameter estimation, two options are available for estimating (or smoothing) the state values. One possibility, is to use the iPDA parameter estimates, $\hat{\boldsymbol{\theta}}$, in the model and solve the differential equations numerically, ignoring the stochastic disturbance. The other way is to use $x_\sim(t)$ as a state estimate (like a Kalman smoother) as is shown in Fig. 4.

To compare the parameter estimates obtained using iPDA and traditional NLS, and to examine the effect of the iPDA weighting factor on the quality of the parameter estimates, parameters k_{ref} and E/R were estimated using iPDA with three different weighting factors ($0.1\lambda_{opt}$, λ_{opt}, $10\lambda_{opt}$) as well as traditional NLS, for 50 simulated data sets. The initial parameter guesses were 50% of the true parameter values. iPDA (NLS) iterations were stopped, when the change in $||\boldsymbol{\tau}||$ ($||\boldsymbol{\theta}||$ in case of NLS) was less than 1E−12, or if a maximum of 1000 iterations was reached. Figs. 6 and 7 show the corresponding box plots for the estimates of k_{ref} and E/R, respectively. Both iPDA and traditional NLS produced good and comparable estimates. iPDA$_{opt}$ (iPDA with λ_{opt}) gave the most precise estimates.

When $\lambda = 10\lambda_{opt}$ was used, the iPDA parameter estimates were closer to those obtained traditional NLS than the estimates using $\lambda = \lambda_{opt}$. This result was expected, because traditional NLS assumes that $\eta(t)$ can be neglected because $\sigma_p^2 \to 0$, which corresponds to very large λ. When we attempted iPDA estimation

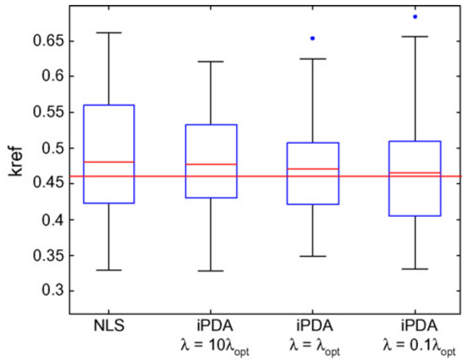

Fig. 6. Box-plots for k_{ref} using NLS and iPDA linear SISO CSTR.

with very large weighting factors (e.g., $\lambda_{opt-discrete} = 100\,min^2$) the iterations stopped before reaching the NLS parameter estimates. Unfortunately, the spline knot sequence that we specified was too coarse for the empirical B-spline curve to be able to solve the ODE arbitrarily well. As a result, the PEN term $\lambda\int((dC'_{A\sim}(t)/dt) - w_C C'_{A\sim}(t) - w_T T'(t))^2\,dt$ in the iPDA objective function was not able to approach zero, and remained large, relative to the SSE term. We anticipate that a large number of spline knots (and perhaps long computational times) will be required to obtain accurate iPDA parameter estimates using very large values of λ. This is not a serious problem for modelers, because in situations where a large λ is appropriate, traditional NLS or collocation-based methods provide good parameter estimates, and iPDA would not be particularly beneficial. We advocate that iPDA be used in situations where there are significant process disturbances, uncertainties in input variables, or an imperfect or simplified ODE model, so that traditional least-squares assumptions do not apply. In such situations, very large values of λ are not required. To confirm that we had used a sufficient number of spline knots to obtain good iPDA parameter estimates for our stochastic CSTR problem, we re-estimated the parameters using twice as many spline knots and obtained almost exactly the same results.

To confirm that iPDA becomes even more beneficial when there are significant process disturbances, we generated additional simulated data sets using the same measurement noise variance as in Fig. 4, but 10 times the amount of process noise variance (making the optimal choice of λ 10 times smaller). As expected, because of the larger process noise, both iPDA and traditional NLS parameter estimates worsened, but the traditional NLS parameter estimates deteriorated more than iPDA parameter estimates.

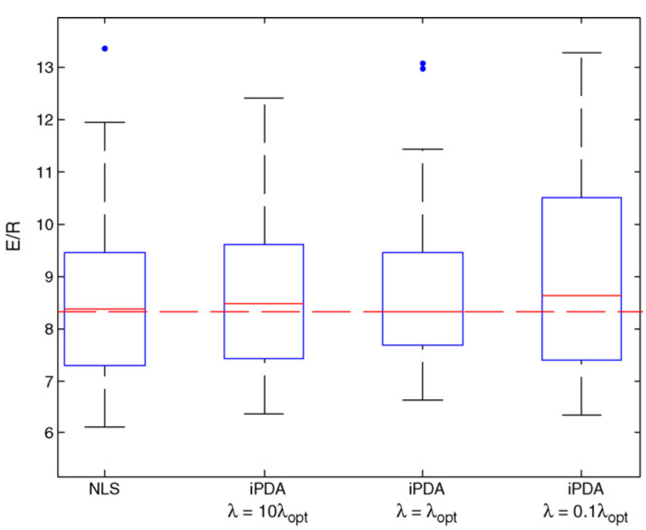

Fig. 7. Box-plots for E/R using NLS and iPDA, linear SISO CSTR.

3.2. Nonlinear MIMO CSTR

The second case study is a non-isothermal CSTR. The model equations consist of material and energy balances (Marlin, 2000) with additional stochastic terms:

$$
\begin{aligned}
\frac{dC_A(t)}{dt} &= \frac{F(t)}{V}(C_{A0}(t) - C_A(t)) - g(T(t))C_A(t) + \eta_1(t)\\
\frac{dT(t)}{dt} &= \frac{F(t)}{V}(T_0(t) - T(t)) + \beta_1(T(t) - T_{cin}(t)) + \beta_2 g(T(t))C_A(t) + \eta_2(t)\\
C_A(0) &= 1.569\,(kmol\,m^{-3})\\
T(0) &= 341.37\,(K)\\
y_1(t_i) &= C_A(t_i) + \varepsilon_1(t_i)\\
y_2(t_j) &= C_A(t_j) + \varepsilon_2(t_j)\\
g(T) &= k_{ref}\exp\left(-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right),\\
\beta_1(F_c) &= -\frac{aF_c^{b+1}}{V\rho C_p(F_c + (aF_c^b/2\rho_c C_{pc}))},\ \beta_2 = \frac{-\Delta H_{rxn}}{\rho C_p}
\end{aligned}
\tag{34}
$$

$E\{\eta_1(t_i)\eta_1(t_j)\} = \sigma_{p1}^2\delta(t_i - t_j)$, $E\{\eta_2(t_i)\eta_2(t_j)\} = \sigma_{p2}^2\delta(t_i - t_j)$, $\varepsilon_1(t_{mj})j = 1,\ldots,N_1$ and $\varepsilon_2(t_{mj})j = 1,\ldots,N_2$ are white-noise sequences with variances σ_{m1}^2 and σ_{m2}^2, respectively. We also assume that η_1, η_2, ε_1, and ε_2 are independent.

This stochastic differential equation model is nonlinear in the states (C_A, T) and parameters and does not have an analytical solution.

As in the previous example, C_A is the concentration of the reactant A, T is the reactor temperature, V is the volume and $T_{ref} = 350\,K$ is the reference temperature. The parameters to be estimated and their true values are the same as those of the SISO case study: $E/R = 8330.1\,K$, $k_{ref} = 0.461\,min^{-1}$. The initial parameter guesses were set at 50% of the true parameter values. This nonlinear system has five inputs: the reactant flow rate F, the inlet reactant concentration C_{A0}, the inlet temperature T_0, the coolant inlet temperature, and the coolant flow rate F_c. Parameters $a = 1.678E6$, $b = 0.5$ which we assume to be known from previous heat-transfer experiments, account for the effect of F_c on the heat-transfer coefficient. Values for the various other known constants (Marlin, 2000) are as follows: $V = 1.0\,m^3$, $C_p = 1\,cal\,g^{-1}\,K^{-1}$, $\rho = 1E6\,g\,m^{-3}$, $C_{pc} = 1\,cal\,g^{-1}\,K^{-1}$, $\rho_c = 1E6\,g\,m^{-3}$, and $-\Delta H_{rxn} = 130E6\,cal\,kmol^{-1}$. The initial steady-state operating point is: $C_{As} = 1.569\,kmol\,m^{-3}$ and $T_s = 341.37\,K$. The linear SISO case study used a simplified form of this model that assumed perfect temperature control.

In this example, there is no temperature controller, and perturbations are introduced into each of the five inputs using the input scheme shown in Fig. 8 (Poyton, 2005). Each input consists of a step up, followed by a step down back to the steady-state point.

We assume that concentration can be measured once per minute and temperature can be measured once every 0.3 min. The duration of the simulation is 64 min, so that there are 64 concentration measurements and 213 temperature measurements. The noise variances for the concentration and temperature measurements are $\sigma_{m1}^2 = 4\times10^{-4}\,(kmol/m^3)^2$ and $\sigma_{m2}^2 = 6.4\times10^{-1}\,K^2$, respectively. The corresponding process noise intensities are $Q_{p1} = 4\times10^{-3}$ $(kmol/m^3)^2/min$ and $Q_{p2} = 4\,K^2/min$. Since $\sigma_{p1}^2 = Q_{p1}/\Delta t$ and $\sigma_{p2}^2 = Q_{p2}/\Delta t$, for $\Delta t = 1\,min$ the corresponding process noise variances are $\sigma_{p1}^2 = 4\times10^{-3}\,(kmol/m^3/min)^2$ and $\sigma_{p2}^2 = 4\,(K/min)^2$.

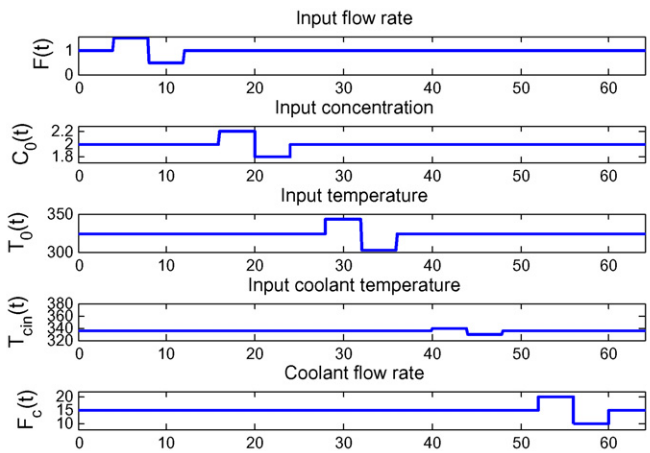

Fig. 8. Input scheme for MIMO nonlinear CSTR.

From Eq. (29), the iPDA objective function is:

$$\sum_{j=1}^{64}(C_{\rm A}(t_{m1j})-C_{\rm A\sim}(t_{m1j}))^2 + \lambda_1\int_{t=0}^{64}\left(\frac{{\rm d}C_{\rm A\sim}(t)}{{\rm d}t}-\frac{F(t)}{V}(C_{\rm A0}(t)-C_{\rm A\sim}(t))+g(T_\sim(t))C_{\rm A\sim}(t)\right)^2{\rm d}t + \sum_{j=1}^{213}(T(t_{m2j})-T_\sim(t_{m2j}))^2 + \lambda_2\int_{t=0}^{64}\left(\frac{{\rm d}T_\sim(t)}{{\rm d}t}-\frac{F(t)}{V}(T_0(t)-T_\sim(t))-\beta_1(T_\sim(t)-T_{\rm cin}(t))-\beta_2 g(T_\sim(t))C_{\rm A\sim}(t)\right)^2{\rm d}t \quad (35)$$

From (11), the optimal weighting factors in this case are $\lambda_1 = 0.1$ min and $\lambda_2 = 0.16$ min. We used two different knot sequences to fit B-spline curves to concentration and temperature observations, with the knots placed at observation times. Again, we obtained optimal parameter values using the iterative procedure shown in Fig. 1 iPDA and also by simultaneously estimating the B-spline coefficients along with the two parameters in Eq. (35) using "lsqnonlin" routine in Matlab. However, in this example the simultaneous approach proved to be much slower (due to the existence of two states and considerably fewer concentration observations), so that the iterative approach was preferred. To compare the sampling behaviour of the iPDA and NLS parameter estimates, 50 sets of concentration and temperature measurements were generated using different measurement and process noise sequences. Gaussian quadrature was used to calculate the integrals in Eq. (33). Box-plots for the parameter estimates are shown in Figs. 9 and 10. Both iPDA and traditional NLS produced reasonable estimates; the iPDA parameter estimates are better, on average, than the traditional NLS estimates.

The iPDA and NLS predicted responses are compared against the true responses in Figs. 11–14. The iPDA predicted responses closely follow the true trajectories.

4. Summary and conclusions

Parameters were estimated in differential equation models with stochastic disturbances using iPDA. By considering the joint probability density of the states and observations, given the parameters, we demonstrated that optimal model parameters and B-spline coefficients can be obtained by minimizing the iPDA objective function. We also demonstrated that the optimal value of the

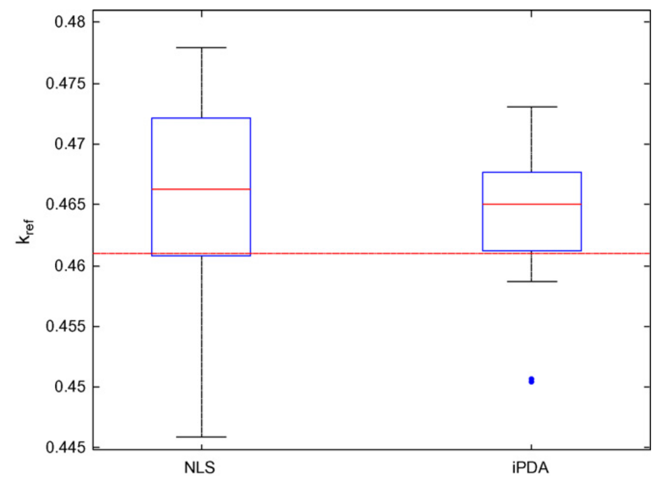

Fig. 9. Box-plots for $k_{\rm ref}$ using NLS and iPDA nonlinear MIMO CSTR.

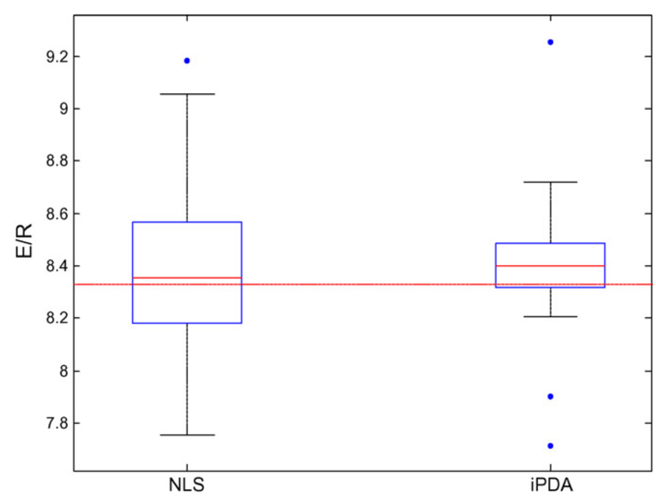

Fig. 10. Box-plots for E/R using NLS and iPDA nonlinear MIMO CSTR.

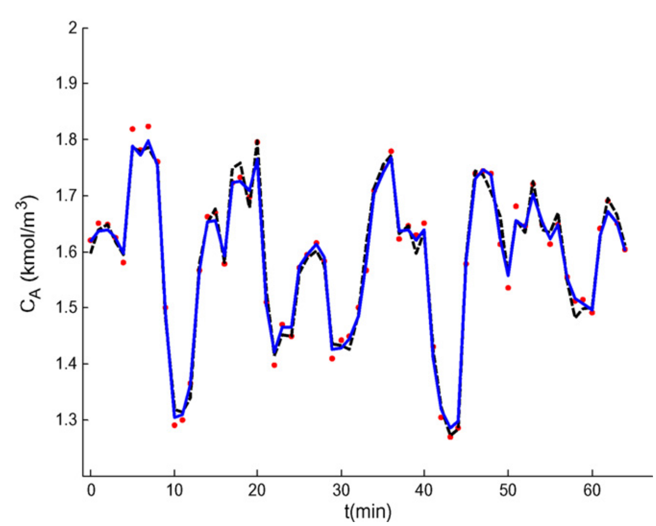

Fig. 11. Observed, true, and predicted concentration response for iPDA for the nonlinear MIMO CSTR example (●, simulated data; - - -, response of the system with true parameters and true stochastic noise; —, iPDA response).

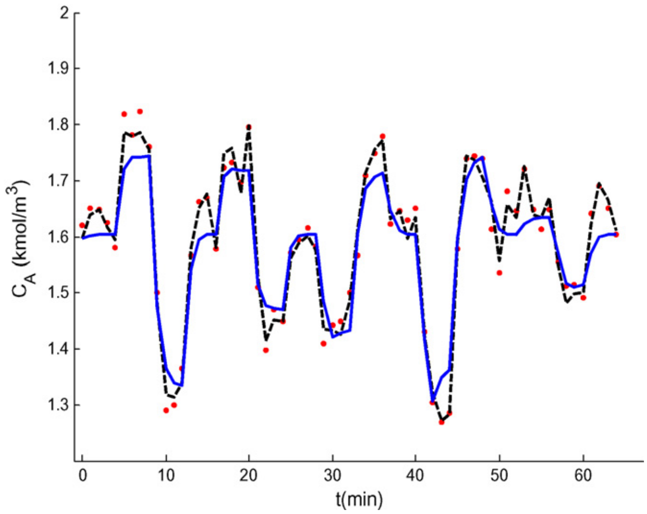

Fig. 12. Measured, true, and predicted concentration responses using NLS for the nonlinear MIMO CSTR example (●, simulated data; - - -, response of the system with true parameters and true stochastic noise; —, NLS response).

weighting factor λ is proportional to the measurement noise variance, and inversely proportional to the model disturbance variance, so that tuning the weighting factor in iPDA resembles tuning the Kalman gain in Kalman filtering applications. For parameter estimation in MIMO dynamic models, the overall iPDA objective function includes a sum-of-squared errors term for each response, with the reciprocal of the corresponding measurement variance as a weighting factor, and a model-based penalty term for each differential equation, with the reciprocal of the process noise variance as a weighting factor. Optimizing the iPDA objective function, using either a simple iterative two-step procedure (in which spline coefficients are estimated in the first step and fundamental parameters are estimated in the second) or simultaneous estimation of spline coefficients and model parameters, produces approximate maximum likelihood estimates for the parameters, conditional on the data and knowledge about the measurement and disturbance variances.

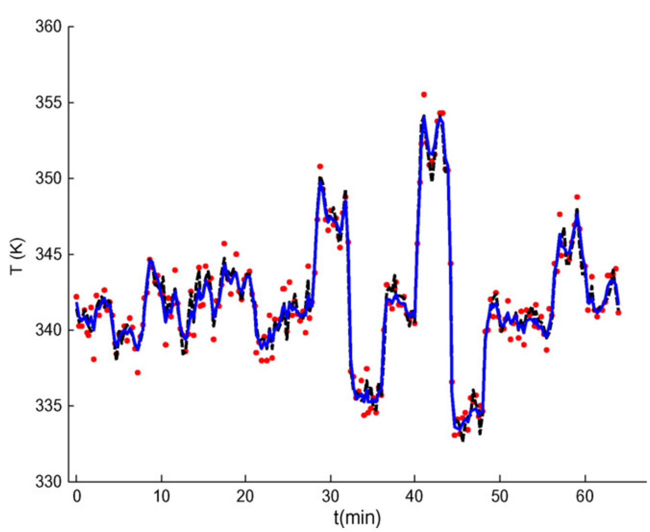

Fig. 13. Observed, true, and predicted temperature response for iPDA for the nonlinear MIMO CSTR example (●, simulated data; - - -, response of the system with true parameters and true stochastic noise; —, iPDA response).

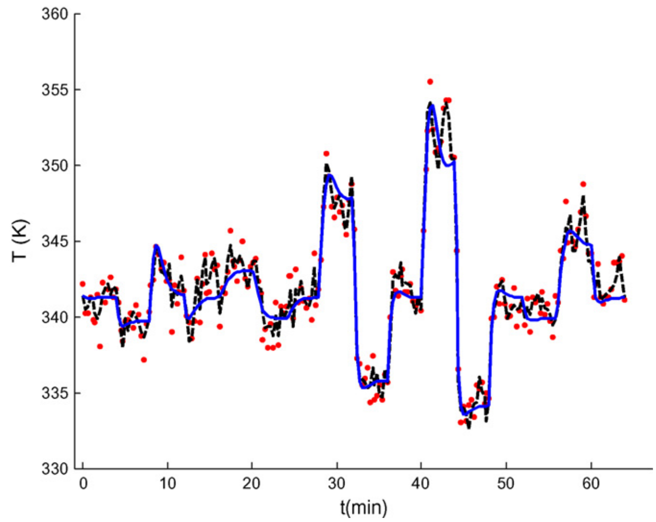

Fig. 14. Observed, true, and predicted temperature response for NLS for the nonlinear MIMO CSTR example (●, simulated data; - - -, response of the system with true parameters and true stochastic noise; —, NLS response).

Two examples were used to validate the results. In the first, parameters were estimated in a linearized SISO differential equation model with a stochastic disturbance using both traditional NLS and iPDA. iPDA parameter estimates were superior to those obtained using traditional NLS, and iPDA was effective in reconstructing the true underlying response trajectory. When iPDA was used with λ larger than the optimal value, parameter estimates were similar to those obtained using traditional NLS because large values of λ are appropriate when the traditional NLS assumptions (i.e., negligible stochastic disturbance compared to the measurement noise) hold. In the second case study, two parameters in a nonlinear MIMO CSTR with a stochastic disturbance were estimated. Similar to the first case study, iPDA performed better than traditional NLS in estimating parameters and reconstructing the response trajectory.

From the case studies we see that iPDA (like a Kalman filter) can be used as a state smoother. Recent simulation studies (Varziri, 2008) have confirmed that iPDA can also be used to observe unmeasured states and to estimate parameters when some of the states are unmeasured.

In our case studies, the measurement noise variance and the model disturbance variance were assumed known. In practical situations however, this is not the case. Although knowledge about measurement variances can be obtained from replicate measurements, it is difficult to obtain a priori knowledge about the model disturbance variance. One objective of our ongoing work is to establish a means of estimating the weighting factor without the limiting assumption of known variances. We are also working to obtain confidence interval expressions for the model parameters.

In iPDA the objective function is minimized by an iterative approach but other approaches for minimizing the same objective function are also possible. In particular, nonlinear and quadratic programming techniques that can handle large-scale minimization problems are potential candidates. These methods have been applied to problems in which orthogonal collocation on finite elements are used to discretize the ODEs (Biegler & Grossman, 2004; Tanartkit & Biegler, 1995; Tjoa & Biegler, 1991), and we are hopeful that they can be used effectively to reduce the computational requirements of iPDA.

Appendix A

In this section, the results of Section 2 are repeated for a multivariate case.

We consider a multi-input multi-output first order, nonlinear system defined as follows:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \\ \mathbf{x}(t_0) &= \mathbf{x}_0 \\ \mathbf{y}(t_j) &= \mathbf{C}\mathbf{x}(t_j) + \boldsymbol{\varepsilon}(t_j) \end{aligned} \quad (36)$$

where $\mathbf{x}$ is the vector of n state variables, $\mathbf{F}$ (nonlinear function) is an $n \times 1$ vector, $\mathbf{u}$ is the vector of u input variables, $\mathbf{y}$ is the vector of m output variables, and $\mathbf{C}$ is a $m \times n$ matrix. $\mathbf{x}_0$ is a multivariate normal random variable with mean $E\{\mathbf{x}_0\}$ and $\mathrm{cov}(\mathbf{x}_0) = \boldsymbol{\Sigma}_0$.

$\boldsymbol{\eta}(t)$ is a continuous zero-mean stationary white noise process with covariance matrix $E\{\boldsymbol{\eta}(t_1)\boldsymbol{\eta}(t_2)^T\} = \mathbf{Q}\delta(t_2 - t_1)$, where $\mathbf{Q}$ is the corresponding power spectral density and $\delta(.)$ is the Dirac delta function. For the discrete time white noise process:

$$E\left\{\boldsymbol{\eta}(j_1\ \Delta t)\boldsymbol{\eta}^T(j_2\,\Delta t)\right\} = \begin{cases} \dfrac{\mathbf{Q}}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \quad (37)$$

where j_1 and j_2 are integers and Δt is the sampling period. $\boldsymbol{\varepsilon}(t_j) = [\varepsilon_{1j}, \ldots, \varepsilon_{mj}]^T$ is a vector of m zero-mean random variables. We assume that measurements of different responses from the same experimental run are independent.

$$E\{\boldsymbol{\varepsilon}(t_i)\boldsymbol{\varepsilon}^T(t_j)\} = \begin{cases} \mathbf{E} & i = j \\ 0 & i \neq j \end{cases} \quad (38)$$

$\mathbf{E}$ is diagonal with $\mathrm{diag}(\mathbf{E}) = [\sigma_{1j}^2, \ldots, \sigma_{mj}^2]^T$ for the jth run. We also assume that measurements from different experimental runs are independent. Whenever the above assumptions are not appropriate, they can be made more general. However, the results in that case maybe more complex depending on the covariance matrix structure.

We assume that the response of the above stochastic system can be approximated by a linear combination of some B-splines:

$$x_{i\sim}(t) = \varphi_i^T(t)\boldsymbol{\beta}_i \text{ for } i = 1, \ldots, n \quad (39)$$

where $\varphi_i(t)$ is a vector containing c_i basis functions, $\boldsymbol{\beta}_i$ is vector of c_i spline coefficients and $\omega_i(t)$ is the stochastic term. Therefore

$$\mathbf{x}_\sim(t) = \begin{bmatrix} \varphi_1^T(t) & 0 & \cdots & 0 \\ 0 & \varphi_2^T(t) & \cdots & 0 \\ 0 & 0 & \cdots & \varphi_n^T(t) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix} = \boldsymbol{\Phi}(t)\boldsymbol{\beta}_c \quad (40)$$

where

$$\boldsymbol{\Phi}(t) = \begin{bmatrix} \varphi_1^T(t) & 0 & \cdots & 0 \\ 0 & \varphi_2^T(t) & \cdots & 0 \\ 0 & 0 & \cdots & \varphi_n^T(t) \end{bmatrix} \text{ and } \boldsymbol{\beta}_c = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix}$$

is the concatenated vector of spline coefficients.

Again we consider Eq. (13), from the same arguments that were used for the SISO case, we will get the same expression for $p(\mathbf{x}_0, \ldots, \mathbf{x}_q|\boldsymbol{\theta})$ as before:

$$p(x_0, \ldots, x_q|\boldsymbol{\theta}) \propto \exp\left\{-\int (\dot{\mathbf{x}}(t) - \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^T \mathbf{Q}^{-1}(\dot{\mathbf{x}}(t) - \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))\,\mathrm{d}t\right\}$$
$$\times \exp\{-(\mathbf{x}_0 - E\{\mathbf{x}_0|\boldsymbol{\theta}\})^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_0 - E\{\mathbf{x}_0|\boldsymbol{\theta}\})\} \quad (41)$$

Suppose the ith response ($i = 1, \ldots, m$) is measured N_i times observations. Let $\mathbf{y}_c = [y_{11} \cdots y_{1N_1} \cdots y_{m1} \cdots y_{mN_m}]^T$ and $\boldsymbol{\varepsilon}_c = [\varepsilon_{11} \cdots \varepsilon_{1N_1} \cdots \varepsilon_{m1} \cdots \varepsilon_{mN_m}]^T$ is zero mean with $\mathrm{cov}(\boldsymbol{\varepsilon}_c) = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{N_1} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \sigma_m^2 \mathbf{I}_{N_m} \end{bmatrix}$. Considering $\mathbf{x}_c = [x_{11} \cdots x_{1N_1} \cdots x_{m1} \cdots x_{mN_m}]^T$

$$\mathbf{y}_c = \mathbf{C}_c\mathbf{x}_c + \boldsymbol{\varepsilon}_c \quad (42)$$

$$E\{\mathbf{y}_c|\mathbf{x}_c, \boldsymbol{\theta}\} = \mathbf{x}_c \quad \text{and} \quad \mathrm{cov}\{\mathbf{y}_c|\mathbf{x}_c, \boldsymbol{\theta}\} = \boldsymbol{\Sigma} \quad (43)$$

where $\mathbf{C}_c = \begin{bmatrix} c_{11}\mathbf{I}_{N_1} & \cdots & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \cdots & c_{mm}\mathbf{I}_{N_m} \end{bmatrix}$. From (43):

$$p(\mathbf{y}_c|\mathbf{x}_0, \ldots, \mathbf{x}_q, \boldsymbol{\theta}) \propto \exp\{-(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_c)^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_c)\} \quad (44)$$

From (13), (41), and (44):

$$p(\mathbf{x}_0, \ldots, \mathbf{x}_q, \mathbf{y}_c|\boldsymbol{\theta}) = \exp\{-(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_c)^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_c)\}$$
$$\times \exp\left\{-\int (\dot{\mathbf{x}}(t) - \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^T \mathbf{Q}^{-1}(\dot{\mathbf{x}}(t) - \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))\,\mathrm{d}t\right\}$$
$$\times \exp\{-(\mathbf{x}_0 - E\{\mathbf{x}_0|\boldsymbol{\theta}\})^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_0 - E\{\mathbf{x}_0|\boldsymbol{\theta}\})\} \quad (45)$$

Therefore the optimal state and parameter estimates, $\hat{\mathbf{x}}$, $\hat{\boldsymbol{\theta}}$ minimize

$$(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_c)^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_c)$$
$$+ \int (\dot{\mathbf{x}}(t) - \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^T \mathbf{Q}^{-1}(\dot{\mathbf{x}}(t) - \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))\ \mathrm{d}t$$
$$+ (\mathbf{x}_0 - E(\mathbf{x}_0|\boldsymbol{\theta}))^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_0 - E\{\mathbf{x}_0|\boldsymbol{\theta}\}) \quad (46)$$

If we assume that the initial condition is perfectly known and that the states can be approximated by Eq. (40) we have:

$$(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c})$$
$$+ \int (\dot{\mathbf{x}}_\sim(t) - \mathbf{F}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}))^T \mathbf{Q}^{-1}(\dot{\mathbf{x}}_\sim(t) - \mathbf{F}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}))\ \mathrm{d}t \quad (47)$$

If the integral in (47) is discretized:

$$(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c})$$
$$+ \sum_{i=0}^{q} (\dot{\mathbf{x}}_\sim(t_i) - \mathbf{F}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}))^T \boldsymbol{\Sigma}_p^{-1}(\dot{\mathbf{x}}_\sim(t_i) - \mathbf{F}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta})) \quad (48)$$

where $\boldsymbol{\Sigma}_p = \mathbf{Q}/\Delta t$.

References

Ascher, U. M., Mathheij, R. M. M., & Russell, R. D. (1988). *Numerical solutions of boundary value problems for ordinary differential equations*. New Jersey: Prentice Hall.

Astrom, K. J. (1970). *Introduction to stochastic control theory*. Academic Press.

Bard, Y. (1974). *Nonlinear parameter estimation*. New York: Academic Press, Inc.

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. John Wiley and Sons, Inc.

Benson, M. (1979). Parameter fitting in dynamic models. *Ecological Modelling*, *6*, 97–115.

Biegler, L. T. (1984). Short note solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers & Chemical Engineering*, *8*, 243–248.

Biegler, L. T., & Grossman, I. E. (2004). Retrospective on optimization. *Computers & Chemical Engineering*, *28*, 1169–1192.

Binder, T., Blank, L., Dahmen, W., & Marquardt, W. (2002). On the regularization of dynamic data reconciliation problems. *Journal of Process Control*, *12*, 557–567.

Bock, H. G. (1981). Numerical treatment of inverse problems in chemical reaction kinetics. In K. H. Ebert, P. Deuflhard, & W. Jäger (Eds.), *Modelling of chemical reaction systems* (pp. 102–125). Springer (Chapter 8).

Bock, H. G. (1983). Recent advances in parameter identification for ordinary differential equations. In P. Deuflhard & E. Hairer (Eds.), *Progress in scientific computing* (pp. 95–121). Birkhäuser.

Brown, R. G., & Hwang, P. Y. C. (1992). *Introduction to random signals and applied Kalman filtering* (2nd ed.). John Wiley & Sons, Inc.

de Boor, C. (2001). *A practical guide to splines*. New York: Springer.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, *11*(2), 89–121.

Gagnon, L., & MacGregor, J. F. (1991). State estimation for continuous emulsion polymerization. *Canadian Journal of Chemical Engineering*, *69*, 648–656.

Gong, J., Wahba, G., Johnson, D. R., & Tribbia, J. (1998). Adaptive tuning of numerical weather prediction models: Simultaneous estimation of weighting, smoothing, and physical parameters. *Monthly Weather Review*, *126*, 210–231.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press.

Kirsch, A. (1996). *An introduction to the mathematical theory of inverse problems*. Springer.

Leis, J. R., & Kramer, M. A. (1988). ALGORITHM 658: ODESSA—An ordinary differential equation solver with explicit simultaneous sensitivity analysis. *ACM Transactions on Mathematical Software*, *14*, 61–67.

Li, Z., Osborne, M. R., & Pravan, T. (2005). Parameter estimation for ordinary differential equations. *IMA Journal of Numerical Science*, *25*, 264–285.

Logsdon, J. S., & Biegler, L. T. (1992). Decomposition strategies for large-scale dynamic optimization problems. *Chemical Engineering Science*, *47*, 851–864.

Lucero, J. C. (2002). Identifying a differential equation for lip motion. *Medical Engineering and Physics*, *24*, 521–528.

Marlin, T. E. (2000). *Process control: Designing processes and control systems for dynamic performance* (2nd ed.). McGraw-Hill.

Maybeck, P. S. (1979). *Stochastic models, estimation, and control* Academic Press.

Ogunnaike, B. A., & Ray, W. H. (1994). *Process dynamics, modeling and control*. New York: Oxford University Press.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, *1*(4), 502–518.

Poyton, A. (2005). *Application of principal differential analysis to parameter estimation in fundamental dynamic models*. M.Sc. Thesis. Kingston, Canada: Queen's University.

Poyton, A. A., Varziri, M. S., McAuley, K. B., McLellan, P. J., & Ramsay, J. O. (2006). Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & Chemical Engineering*, *30*, 698–708.

Ramsay, J. O. (1996). Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society Series B (Methodological)*, *58*, 495–508.

Ramsay, J. O. (2000). Functional components of variation in handwriting. *Journal of the American Statistical Association*, *95*, 9–16.

Ramsay, J. O., & Munhall, K. G. (1996). Functional data analyses of lip motion. *Journal of the Acoustical Society of America*, *99*, 3718–3727.

Ramsay, J. O., & Ramsey, J. B. (2002). Functional data analysis of the dynamics of the monthly index of nondurable goods production. *Journal of Econometrics*, *107*, 327–344.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). Springer.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. John Wiley and Sons, Inc.

Swartz, J., & Bremermann, H. (1975). Discussion of parameter estimation in biological modelling: Algorithms for estimation and evaluation of estimates. *Journal of Mathematical Biology*, *1*, 241–275.

Tanartkit, P., & Biegler, L. T. (1995). Stable decomposition for dynamic optimization. *Industrial Engineering Chemistry Research*, *34*, 1253–1266.

Tang, Y. P. (1971). On the estimation of rate constants for complex kinetic models. *Industrial Engineering Chemistry Fundamentals*, *10*, 321–322.

Tjoa, I.-B., & Biegler, L. T. (1991). Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems. *Industrial Engineering Chemistry Research*, *30*, 376–385.

Vajda, S., & Valko, P. (1986). A direct–indirect procedure for estimation of kinetic parameters. *Computers & Chemical Engineering*, *10*(1), 49–58.

Varah, J. M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific Computing*, *3*, 28–46.

Varziri, M. S. (2008). *Parameter estimation in nonlinear continuous-time dynamic models with modeling errors and process disturbances*. Ph.D. Thesis. Queen's University, in preparation.

Varziri, M. S., McAuley, K. B., & McLellan, P. J. (2008). Parameter estimation in continuous-time dynamic models in presence of unmeasured states and nonstationary disturbances. *Industrial & Engineering Chemistry Research*, *47*, 380–393.

Varziri, M. S., McAuley, K. B., & McLellan, P. J. (submitted for publication-a). Parameter estimation in nonlinear stochastic continuous-time dynamic models with unknown disturbance intensity. *Canadian Journal of Chemical Engineering*.

Varziri, M. S., McAuley, K. B., & McLellan, P. J. (submitted for publication-b). Approximate maximum likelihood parameter estimation for nonlinear dynamic models: Application to a lab-scale nylon reactor model. *Industrial & Engineering Chemistry Research*.