Inference in Dynamic Error-in-Variable-Measurement Problems

João S. Albuquerque and Lorenz T. Biegler
Dept. of Chemical Engineering

Robert E. Kass
Dept. of Statistics

Carnegie Mellon University, Pittsburgh, PA 15213

Efficient algorithms were developed for estimating model parameters from measured data, even in the presence of gross errors. In addition to point estimates of parameters, however, assessments of uncertainty are needed. Linear approximations provide standard errors, but they can be misleading when applied to models that are substantially nonlinear. To overcome this difficulty, profiling methods were developed for the case in which the regressor variables are error-free. These methods provide accurate nonlinear confidence regions, but become expensive for a large number of parameters. These profiling methods are modified to error-in-variable-measurement models with many incidental parameters. Laplace's method is used to integrate out the incidental parameters associated with the measurement errors, and then profiling methods are applied to obtain approximate confidence contours for the parameters. This approach is computationally efficient, requires few function evaluations, and can be applied to large-scale problems. It is useful when certain measurement errors (such as input variables) are relatively small, but not so small that they can be ignored.

Introduction

Parameter estimation, data reconciliation, and gross error detection for error-in-variable measurements (EVM) cases of both steady-state and dynamic systems have been treated extensively in the literature (Albuquerque and Biegler, 1996a,b; Basu and Paliwal, 1989; Crowe, 1989, Fariss and Law, 1979; Johnston and Kramer, 1995; Kretsolvalis and Mah, 1988; Liebman et al., 1992; Narasimhan and Mah, 1988; Sistu et al., 1993; Swartz, 1989; Tamhane et al., 1992; Tjoa and Biegler, 1991; Schlöder et al., 1995) leading to algorithms that are able to solve many large optimization problems that arise in common industrial practice. On the other hand, little attention has been given to the important concomitant problem of assessing uncertainty in these estimates, and drawing statistical inferences on the parameters. In problems that are approximately linear, standard errors may be obtained by applying standard least-squares theory to a linearized solution space for the system. As nonlinearity increases, however, alternative methods are needed. An effective approach involves profiling (Bates and Watts, 1988), which is a technique based on the signed-square root of the loglikelihood function (or log posterior density) that produces improved confidence intervals (Sweeting, 1995) on the estimated parameters. The purpose of this article is to extend the profiling methodology to EVM problems. The additional complications in this context are (1) the introduction of incidental parameters associated with the measurement errors in control variables; (2) the computational burden of having to solve the system each time the loglikelihood function is evaluated. Here, we adopt a Bayesian approach and apply Laplace's method to integrate out the incidental parameters. We take advantage of the sparse structure of the problem and find this methodology to be useful when the size of the measurement error is small, but not so small that it may be ignored.

The dynamic EVM parameter estimation problem and the available solution methods for it are described. This problem is encountered frequently in process control applications. Here, both the outputs and the inputs are measured and

Correspondence concerning this article should be addressed to L. T. Biegler.
Current address of J. S. Albuquerque: Aspen Technology, Cambridge, MA 02139.

model parameters are inferred. In this problem the control input variables act as incidental parameters, and they generally have small measurement errors (such as actuator errors) associated with them. The profiling methods are reviewed for drawing inferences on the parameters for problems where the regressor variables are error free. The way that Laplace's method may be used to integrate out the incidental parameters is described, and several examples are presented.

Dynamic EVM: Formulation and Point Estimates

Consider a process whose state depends on the values of input or control variables (such as feed flows and temperatures) and unknown time invariant parameters (such as rates of reaction and heat-transfer coefficients). This process will be described by a system of differential and algebraic equations (DAE) and in most cases by consistent initial conditions

$$f\left(\frac{dx}{dt},x(t),u(t),\theta,t\right)=0 \tag{1}$$

$$x(t_1)=x_1 \tag{2}$$

Here, $x=x(t)$ are the state variables, $u=u(t)$ are the input or control variables, and $\theta=(\theta_1,\ \ldots,\ \theta_m)$ is the m-dimensional model parameter vector. Equations 1 and 2 define an implicit function $x(t)=g(u(t),\theta,t)$. Thus, the degrees-of-freedom correspond to the parameters θ and the input variables u. In EVM problems, both the state (explanatory) and the input (regression) variables are being measured. The measurement error model is

$$\bar{x}(t)=x(t)+\epsilon_x(t) \tag{3}$$

$$\bar{u}(t)=u(t)+\epsilon_u(t) \tag{4}$$

where $\bar{x}(t)$ and $\bar{u}(t)$ are the measurements associated with the state and input variables, respectively, and $\epsilon_x(t)$ and $\epsilon_u(t)$ are their measurement noises. Because the control input variables have errors chiefly associated with actuators and control elements, we assume that $\epsilon_u(t)$ is small.

Note that in practice, it is unusual to measure all the state variables and, instead, one usually measures a set of variables y related to some subset of state variables. However, for the sake of clarity, we will work with the model described by Eqs. 1–4. Extensions to deal only with measured variables can be made following the same analysis.

Let $p_x(\epsilon_x(t))$ and $p_u(\epsilon_u(t))$ be the probability density functions associated with the measurement noises. The problem here is to estimate the parameter vector θ from measurements $\bar{x}(t)$ and $\bar{u}(t)$, and also the densities $p_x(\epsilon_x(t))$ and $p_u(\epsilon_u(t))$. The measurements will be distributed according to $p_x(\bar{x}(t)\,|\,x(t))$ and $p_u(\bar{u}(t)\,|\,u(t))$. Assuming that both noises $\epsilon_x(t)$ and $\epsilon_u(t)$ are independent, then the joint density of $(x(t),u(t))$ will be $p_x(\bar{x}(t)\,|\,x(t))p_u(\bar{u}(t)\,|\,u(t))$. Since $u(t)$ and θ are implicit functions of $x(t)$, the likelihood function can be written as

$$L(u(t),\theta)=p_x(\bar{x}(t)|x(t))p_u(\bar{u}(t)|u(t)). \tag{5}$$

Defining the prior density as $\pi(u(t),\theta)$, the joint posterior density will then be

$$p(u(t),\theta)=L(u(t),\theta)\pi(u(t),\theta). \tag{6}$$

Using the maximum-a-posteriori (MAP) method, the regression problem can be posed then as an optimization problem

$$\min_{\theta,u(t)} -\log\{p(u(t),\theta)\}=-\log\{p_x[\epsilon_x(t)]p_u[\epsilon_u(t)]\pi(u(t),\theta)\}$$

$$\text{s.t.} f.\left[\frac{dx}{dt},x(t),u(t),\theta,t\right]=0 \tag{7}$$

$$x(t_1)=x_1.$$

We assume state and control variables cannot be sampled continuously across time, but at discrete times $(t_1,\ \ldots,\ t_n)$. Also, the state variables are sampled one time instant ahead of the control variables (at t_{n+1}) and we assume that both the noise and prior distributions are independent across time. Then, the posterior distribution density will become

$$p(u_1,\ \cdots,\ u_n,\theta|\bar{x},\bar{u})$$

$$=p_x(\bar{x}_{n+1}|x_{n+1})\prod_{i=1}^{n}p_x(\bar{x}_i|x_i)p_u(\bar{u}_i|u_i)\pi(u_i,\theta) \tag{8}$$

To ensure an accurate representation, we discretize the DAE system (Eqs. 1 and 2) using Implicit Runge Kutta or orthogonal collocation schemes at these sampling points (Albuquerque and Biegler, 1996a; Liebman et al., 1992; Sistu et al., 1993) obtaining the following collocation and continuity equations

$$F_i(x_i,\dot{x}_i,u_i,\theta)=0$$

$$x_{i+1}=H_i(x_i,\dot{x}_i)$$

$$i=1,\ \ldots,\ n \tag{9}$$

where $\dot{x}_i$ are the collocation variables or stage derivatives. Although the DAE does not have to be discretized only at the sampling points (Albuquerque and Biegler 1996a), we assume that the sampling times provide a stable enough discretization grid for simplicity, keeping in mind that this assumption can be easily relaxed. Using the MAP method on the distribution (Eq. 8) constrained to the discretized DAE (Eq. 9), Eq. 7 is transformed into the following discretized nonlinear programming (NLP) formulation

$$\min_{\theta,u_1,\ \cdots\ ,u_n} -\sum_{i=1}^{n}\log[p_x(\bar{x}_i|x_i)p_u(\bar{u}_i|u_i)\pi(u_i,\theta)]$$

$$-\log p_x(\bar{x}_{n+1}|x_{n+1})$$

$$\text{s.t.}\quad F_i(x_i,\dot{x}_i,u_i,\theta)=0$$

$$x_{i+1}=H_i(x_i,\dot{x}_i) \tag{10}$$

$$i=1,\ \cdots,\ n$$

$$x_1=x(t_1)$$

Equation 10 grows in size both in the number of variables and degrees-of-freedom for optimization as the number of time steps or data sets increases, and can easily become too big to solve unless special measures are taken (Albuquerque and Biegler, 1996a; Steinbach, 1995; Betts, 1995; Schlöder et al., 1995). Several efficient methods have been developed that allow for a fast solution of the larger-scale problems that arise in solving Eq. 10. Some of these methods (Albuquerque and Biegler, 1996a; Steinbach, 1995; Schlöder et al., 1995) are tailored to the structure of the first-order optimal conditions arising from Eq. 10 and their computational effort is linear with the number of time steps (or data sets). General large-scale optimization (Betts, 1995) algorithms with sparse algebra reordering algorithms for the first-order optimal conditions can also be used. However, even when an efficient method is used, the problem may still be expensive. Consequently, any method of drawing inferences on the parameters must limit the number of times Eq. 10 is solved.

Profiling Methods

Profiling methods were discussed by Bates and Watts (1988) as an approximate and reasonably efficient method to obtain nonlinear inferences. Additional general discussion and theory has been provided by Sweeting (1995). In this section we review key elements of this methodology.

Consider the regression problem described in the previous section but without the input or control variables (incidental parameters) $u(t)$ or assuming that these are error free, so that they become constants. The DAE that models the process becomes

$$f\left(\frac{\partial x}{\partial t}, x(t), \theta, t\right) = 0 \tag{11}$$

$$x(t_1) = x_1 \tag{12}$$

and the measurement model is

$$\bar{x}(t) = x(t) + \epsilon(t). \tag{13}$$

As in the previous section, we obtain a likelihood function from the distribution of the measurement error $\epsilon(t)$ and introduce a prior distribution on the parameters $\pi(\theta)$. The resulting posterior density function becomes

$$p(\theta|\bar{x}) \propto p(\bar{x}(t)|x(t))\pi(\theta). \tag{14}$$

Sampling $x(t)$ at times $(t_1, \cdots, t_{n+1})$ and discretizing DAE (Eq. 11) at those times, we get the following MAP problem

$$\begin{aligned} \min_{\theta} \; & -\sum_{i=1}^{n+1} \log[p(\bar{x}_i|x_i)\pi(\theta)] \\ \text{s.t.} \quad & F_i(x_i, \dot{x}_i, \theta) = 0 \\ & x_{i+1} = H_i(x_i, \dot{x}_i) \\ & i = 1, \cdots, n \\ & x_1 = x(t_1) \end{aligned} \tag{15}$$

Define the log posterior function as $\ell(\theta) = -\log(p(\theta|\bar{x}))$, where $p(\theta|\bar{x})$ is defined by Eq. 14, and let $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_m)^T$ be the solution to Eq. 15. In addition, let $\hat{\theta}_{-q} = (\hat{\theta}_1, \cdots, \hat{\theta}_{q-1}, \hat{\theta}_{q+1}, \cdots, \hat{\theta}_p)^T$ be the conditional solution to Eq. 15 subject to a fixed value of θ_q.

For linear problems (using appropriately chosen prior distributions), the posterior distribution of θ has a multivariate student's t distribution, which leads to marginal t (or asymptotically z) distributions for each parameter and bivariate t distributions for each pair of parameters. Thus, inferences about individual parameters may be made, and the bivariate densities for parameter pairs have elliptical contours, which are easily drawn. Corresponding to these methods are complementary techniques for nonlinear problems, which are based on what Bates and Watts call *profile t* functions, *profile traces*, and *profile pair sketches*.

First, the profile t, also more informatively called the signed-square-root of the log posterior density (Sweeting, 1995)

$$\tau(\theta_q) = \operatorname{sgn}\left(\theta_q - \hat{\theta}_q\right)\sqrt{2\left(\ell\left(\theta_q, \hat{\theta}_{-q}\right) - \ell(\hat{\theta})\right)} \tag{16}$$

is a symptotically distributed (as $n \to \infty$) as a standard normal distribution to a high order of accuracy (Bates and Watts, 1988; Seber and Wild, 1989; Sweeting, 1995). The term "profile" here refers to the use of the conditional maximum of the log posterior after fixing the value of the component θ_q. This profile $\tau(\theta_q)$ may be displayed as a function of θ_q and, using the approximation $\tau(\theta_q) \sim N(0,1)$, marginal confidence intervals may be obtained from the cumulative distribution values for the asymptotic normal distribution by solving

$$\tau(\theta_q) = \pm z_{1-\alpha/2} \tag{17}$$

for θ_q (effectively, by reading values from the horizontal scale of the plot). This way, we get an approximate $1-\alpha$ marginal confidence interval $[\theta_q^l, \theta_q^u]$. In addition, nonlinearity of this plot indicates nonlinearity of the model and deviation from normality of the marginal posterior on θ_q.

The profile trace plot for parameters θ_p and θ_q is a plot of θ_p vs. $(\hat{\theta}_{-p})_q$ and also θ_q vs. $(\hat{\theta}_{-q})_p$, where $(\hat{\theta}_{-q})_p$ is the pth element of the vector $(\hat{\theta}_{-q})$. This plot gives information on the joint posterior of θ_p and θ_q. If the parameters are highly correlated, the pair of curves will be very close, whereas if the correlation is weak, the profile traces will tend to be perpendicular. The profile traces will be linear if the model is linear in the parameters. In a profile trace plot of θ_q vs. θ_p, the endpoints of the marginal confidence intervals $[\theta_q^l, \theta_q^u]$ and $[\theta_p^l, \theta_p^u]$ correspond to extreme points. Therefore, the tangents to the confidence contour at these points will be vertical or horizontal. These four points plus this information on the tangents can be used to sketch an approximate confidence contour (the profile pair sketch) by interpolation.

From the above properties, the profiling procedure can be summarized with the following steps, as discussed in Bates and Watts (1988). Further details of this procedure and examples on the construction of approximate nonlinear confidence regions using profiling are provided in Bates and Watts (1988).

• For each parameter θ_p, set τ_p according to the desired confidence levels $1-\alpha$ in Eq. 17 and obtain the corresponding value of θ_p by solving Eq. 16.

For all $q \neq p$:

• For a fixed value of θ_p, determine $(\hat{\theta}_{-p})_q$ by solving Eq. 15 and plot θ_p vs. $(\hat{\theta}_{-p})_q$. From the profile trace plots, we can get the following pairs from the upper and lower limits of θ_p and θ_q

$$\left(\theta_q^l\left(\hat{\theta}_{-q}\right)_p\right) \quad \left(\theta_q^u,\left(\hat{\theta}_{-q}\right)_p\right) \quad \left(\left(\hat{\theta}_{-p}\right)_q,\theta_p^l\right) \quad \left(\left(\hat{\theta}_{-p}\right)_q,\theta_p^u\right)$$

• Transform the values of θ_p vs. $(\hat{\theta}_{-p})_q$ vs. θ_q for each pair into the space of τ_p and τ_q

$$\left(\tau_q^l,\tau_p\left(\tau_q^l\right)\right) \quad \left(\tau_q^u,\tau_p(\tau_q^u)\right) \quad \left(\tau_q^l\left(\tau_p^l\right),\tau_p^l\right) \quad \left(\tau_q^u(\tau_p^u),\tau_p^u\right)$$

• Normalize these points in the τ space by dividing them by the critical confidence value $\tau_{1-\alpha/2}$. Let $\tau' = \tau/\tau_{1-\alpha/2}$ so that we get the following pairs

$$\begin{bmatrix} -1 & \tau_p'(-1) \\ 1 & \tau_p'(1) \\ \tau_q'(-1) & -1 \\ \tau_q'(1) & 1 \end{bmatrix} \tag{18}$$

• For a given confidence level $1-\alpha$, interpolate the points from the pairwise plot in τ-space by approximating the confidence region as an ellipse. This can be done by setting

$$\tau_q' = \cos\left(a + \frac{d}{2}\right) \tag{19}$$

$$\tau_p' = \cos\left(a - \frac{d}{2}\right). \tag{20}$$

Here, the angle a varies from $-\pi$ to π and d is a constant phase shift for a true ellipse. Based on points from the pairwise plots, interpolate d as a function of a, that is

$$\tau_q' = \cos\left(a + \frac{d(a)}{2}\right) \tag{21}$$

$$\tau_p' = \cos\left(a - \frac{d(a)}{2}\right). \tag{22}$$

• For a varying from $-\pi$ to π and $d(a)$, plot the approximate contours in τ-space.

• Transform these contours from τ-space to θ-space.

Integrating Incidental Parameters

The profiling approach described in the previous section cannot be applied directly to Eq. 10 because of the large number of control variables u_i, which are incidental parameters. Within the Bayesian framework we are adopting, the solution is to integrate the incidental parameters u_i from the joint posterior distribution of $(\theta, u_1, \cdots, u_n)$ retrieving the marginal posterior density in θ alone, that is,

$$p(\theta|\bar{x},\bar{u}) \propto \int_{-\infty}^{+\infty} p(\theta,u_1, \cdots, u_n|\bar{x},\bar{u})du_1 \cdots du_n, \tag{23}$$

and then use the profiling techniques outlined above. Because of the heavy computational burden involved in evaluating the posterior density (which requires solving Eq. 10 repeatedly), we use an approximation known as Laplace's method (Kass et al., 1991; Murray, 1974; Tierney and Kadane, 1986). A further justification for this approach is that the error associated with u_i is small.

Laplace's method

Let ϕ be the objective function in the Eq. 10 problem

$$\phi(x_1, \cdots, x_{n+1},u_1, \cdots, u_n,\theta) = \sum_{i=1}^{n} -\log[\, p_x(\bar{x}_i|x_i)p_u(\bar{u}_i|u_i)\pi(u_i,\theta)] - \log\, p_x(\bar{x}_{n+1}|x_{n+1}). \tag{24}$$

For the sake of simplicity, we gather the state and control variables across all data sets into $x^T = (x_1^T, \cdots, x_{n+1}^T)$ and $u^T = (u_1^T, \cdots, u_n^T)$. Note that the state variables x are implicit functions of the parameters θ and the control variables u through the DAE model (Eq. 9); thus, the objective function of the Eq. 10 problem can be written as

$$\Phi(u,\theta) = \phi(x(u,\theta),u,\theta). \tag{25}$$

The posterior density on (u,θ) will be

$$p(u,\theta|\bar{x},\bar{u}) \propto \exp\,(-\Phi(u,\theta)). \tag{26}$$

We expand the objective function around the conditional minimum $(\hat{u},(\theta),\theta)$ in a second-order Taylor series

$$\Phi(u,\theta) \approx \Phi(\hat{u}(\theta),\theta) + \frac{1}{2}(u-\hat{u}(\theta))^T \frac{d^2\Phi(\theta,u)}{du^2}\bigg|_{\hat{u}(\theta),\theta} (u-\hat{u}(\theta)). \tag{27}$$

Inserting Eq. 27 into Eq. 26, we get the following approximation $\hat{p}(\theta|\bar{x},\bar{u})$ to the marginal posterior density $p(\theta|\bar{x},\bar{u})$

$$\hat{p}(\theta|\bar{x},\bar{u}) \propto \exp\,(-\Phi(\hat{u},\theta)) \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}(u-\hat{u})^T \frac{d^2\Phi(\theta,u)}{du^2}\bigg|_{\hat{u},\theta} (u-\hat{u})\right) du. \tag{28}$$

The integral on the righthand side of Eq. 28 is equal to the normalizing constant of a multivariate normal distribution with mean $\hat{u}$ and covariance matrix $[d^2\Phi(\theta,u)/d_u^2|_{\hat{u},\theta}]^{-1}$, thus Eq. 28 becomes

$$\hat{p}(\theta|\bar{x},\bar{u}) \propto \frac{1}{|\Sigma|^{1/2}} \exp(-\Phi(\hat{u},\theta)) \tag{29}$$

where $\Sigma^{-1} = d^2\Phi(\theta,u)/du^2|_{\hat{u},\theta}$. Note that we do not need to compute the proportionality constant in Eq. 29 because the profiling method described in the previous section involves the computation of the difference between log posterior functions (see Eq. 16).

The approximation of Eq. 29 becomes inaccurate when the nonlinearity of the model becomes severe, or when the input variable measurements are very noisy while the number of incidental parameters is large. We will illustrate this effect in the next section of examples.

Computation of the Hessian matrix

We now describe how to compute the Hessian matrix $d^2\Phi(\theta,u)/du^2|_{\hat{u},\theta}$, which is needed for approximation (Eq. 29). This is the matrix of total second derivatives of the objective function (Eq. 25) constrained to the DAE model (Eq. 9) and the initial conditions (Eq. 2), which we write in a compressed form

$$F(x,\dot{x},u,\theta) = \begin{bmatrix} F_1(x_1,\dot{x}_1,u_1,\theta) \\ \vdots \\ F_n(x_n,\dot{x}_n,u_n,\theta) \end{bmatrix} \tag{30}$$

$$G(x,\dot{x}) = \begin{bmatrix} x_1 - x(t_1) \\ x_2 - H_1(x_1,\dot{x}_1) \\ \vdots \\ x_{n+1} - H_n(x_n,\dot{x}_n) \end{bmatrix} \tag{31}$$

where $\dot{x} = (\dot{x}_1^T, \cdots, \dot{x}_n^T)^T$. From Eq. 25, and using the derivative chain rule and the fact that the state variables x are implicit functions of the incidental parameters u, we get the following expression for the total second derivatives

$$\frac{d^2\Phi}{du^2} = \frac{d^2\phi}{du^2} = \frac{\partial^2\phi}{\partial u^2} + \left(\frac{\partial^2\phi}{\partial u \partial x}\right)\frac{dx}{du} + \frac{dx^T}{du}\left(\frac{\partial^2\phi}{\partial x \partial u}\right) + \left(\frac{dx}{du}\right)^T \frac{\partial^2\phi}{\partial x^2}\frac{dx}{du} + \sum_k \frac{d^2x_k}{du^2}\frac{\partial\phi}{\partial x_k}. \tag{32}$$

In order to perform these computations, we will have to compute the first and second derivatives of x with respect to u. These will be computed from Eqs. 30 and 31 using the derivative chain rule. We start by differentiating equations $F(x,\dot{x},u,\theta) = 0$ and $G(x,\dot{x}) = 0$ with respect to x using the derivative chain rule, and then solve for dx/du and d^2x/du^2. For the first derivatives we have

$$\frac{d\dot{x}}{dx} = -\left(\frac{\partial G}{\partial \dot{x}}\right)^{-1}\frac{\partial G}{\partial x} \tag{33}$$

$$\frac{dF}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial \dot{x}}\frac{d\dot{x}}{dx} \tag{34}$$

$$\frac{dx}{du} = -\left(\frac{dF}{dx}\right)^{-1}\frac{\partial F}{\partial u}. \tag{35}$$

For the second derivatives we get

$$\frac{d^2x}{du^2} = -\left(\frac{dF}{dx}\right)^{-1}\left[\frac{\partial^2 F}{\partial u^2} + 2\frac{\partial^2 F}{\partial u \partial x}\frac{dx}{du} + \left(\frac{dx}{du}\right)^T \frac{\partial^2 F}{\partial x^2}\left(\frac{dx}{du}\right)\right]. \tag{36}$$

Note that the model Eqs. 30 and 31 are sparse, and they have an almost diagonal block structure; the derivative of the second entry in Eq. 31 $(x_2 - H(x_1,\dot{x}_1))$ with respect to x_3, for instance, is identically zero, which means that the Eqs. 33, 34, 35, 36 and 32 can be easily tailored to this particular basic structure leading to a computational burden linear with the number of data sets (n). Assuming that the Hessian matrix of the total second derivatives of the objective function with respect to the control variables will be positive definite (in order for the Eq. 10 problem to have a unique solution), computing the determinant of Σ in Eq. 29 will be proportional to n^2 if we use a Cholesky factorization on Σ.

Finally, we note that a less accurate inference based on a linearized model offers no computational advantage for EVM problems. First, we see in the next section that this linear approximation can be very poor. Moreover, since we have a large number of parameters (u,θ), this analysis would be as expensive as the application of Laplace's method. For instance, we could linearize the model (Eqs. 30 and 31) around the optimal point, eliminate the stage derivatives $\dot{x}_i$ and solve for the state variables x_i by replacing the posterior density on (u,θ) (Eq. 26), which can be integrated in u using Eq. 28 to yield a marginal density in θ. Again, for each value of θ we would have to resolve the regression problem to compute the conditional optimal incidental parameters $\hat{u}(\theta)$ in order to capture the confidence regions between u and θ.

Examples

In this section we treat three examples. The first example is very simple so that exact posterior densities can be computed for comparison with approximation (Eq. 29). The second example is nearly linear, and the incidental parameters play a very mild role in the inference process. The third problem has a strong nonlinearity and shows the usefulness of these nonlinear inference techniques, since in this case linear approximations would be quite inaccurate. In all our examples, we use flat prior distributions on the parameter vector θ and the incidental parameters.

Accuracy of Laplace's method in terms of input noise and number of input terms

We illustrate the effect of increasing input noise on approximation (Eq. 29) in a very simple situation. Consider a problem with only one state variable x, one incidental parameter u, and one parameter θ where the measurements are given by

$$\bar{x} = x + \epsilon_x, \quad \epsilon_x \sim N(0,1)$$
$$\bar{u} = u + \epsilon_u, \quad \epsilon_u \sim N(0,\sigma^2) \tag{37}$$

and the state, control and the parameter are related through

$$x = \theta u^2. \tag{38}$$

Assume that the measurements are $\bar{x} = 0$ and $\bar{u} = 1$. Using Eq. 38 and the distributions Eq. 37, we get the following joint probability density for u and θ

$$p(u,\theta|\bar{x},\bar{u}) \propto \frac{1}{\sigma} \exp\left(-\frac{1}{2}\left(\frac{u-1}{\sigma}\right)^2\right)\exp\left(-\frac{1}{2}\theta^2 u^4\right) \tag{39}$$

The marginal density on θ is obtained by integrating out u from Eq. 39

$$p(\theta) = \int_{-\infty}^{\infty} p(u,\theta)du \tag{40}$$

Note that when the measurement noise on the incidental parameters is very small ($\sigma \to 0$), the first exponential on right-hand side of Eq. 30 will tend to a Dirac $\delta(u-1)$ function, and the marginal density on θ will be given by a normal distribution $N(0,1)$. From Eq. 7 we see that this regression problem can be posed as

$$\min_{u,\theta} \frac{1}{2}x^2 + \frac{1}{2}\left(\frac{u-1}{\sigma}\right)^2$$
$$\text{s.t.} x = \theta u^2. \tag{41}$$

When $\sigma \to 0$, the term in u will drop out of the objective function and Eq. 41 becomes a linear parameter estimation problem with $\theta \sim N(0,1)$. By linearizing the constraint in the Eq. 41 problem around the optimal value $\hat{\theta} = 0$, $\hat{u} = 1$, and $\hat{x} = 0$ and replacing in the distributions (Eq. 37) and integrating out the incidental parameter u, we get the following linear approximation for the integral (Eq. 40).

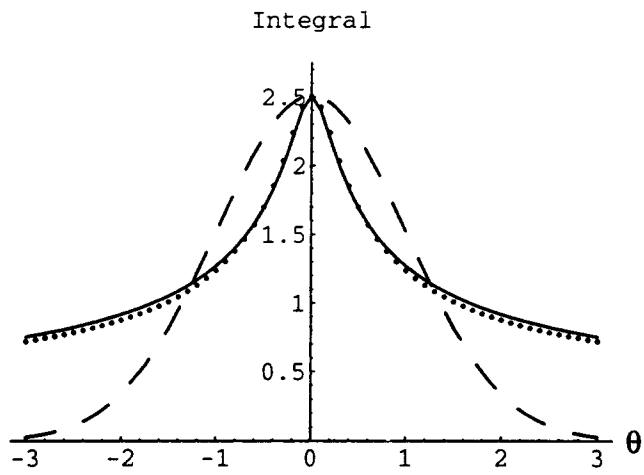

Figure 2. Exact integral on θ (——), Laplace approximation (· · · ·) and linear approximation (----) for noise on incidental parameters $\sigma = 1$.

$$p(\theta|\bar{x},\bar{u}) \approx \sqrt{2\pi}\, \exp\left(-\frac{1}{2}\theta^2\right) \tag{42}$$

which is an unnormalized $N(0,1)$ distribution.

In this example we compare the unnormalized marginal density in θ obtained by computing the integral (Eq. 40) numerically using *Mathematica* (Wolfram, 1988) with the Laplace approximation for several noise levels in u and with the linear approximation (Eq. 42). Figures 1, 2 and 3 show the comparisons for standard-deviation $\sigma = 0.01$, 1.0, and 2.0, respectively. For the first two cases, the Laplace approximation is almost perfect but it deviates for the last case where the incidental parameters become quite noisy and the problem deviates more from linearity. In every case, however, the Laplace approximation is far better than the linear approximation.

We now study the behavior of approximation (Eq. 29) when applied to high-dimensional integrals (large number of inci-

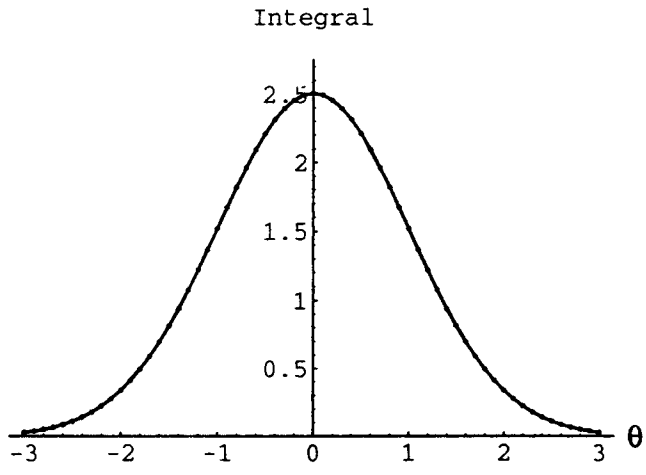

Figure 1. Exact integral on θ (——), Laplace approximation (· · · ·) and linear approximation (----) for noise on incidental parameters $\sigma = 0.01$.

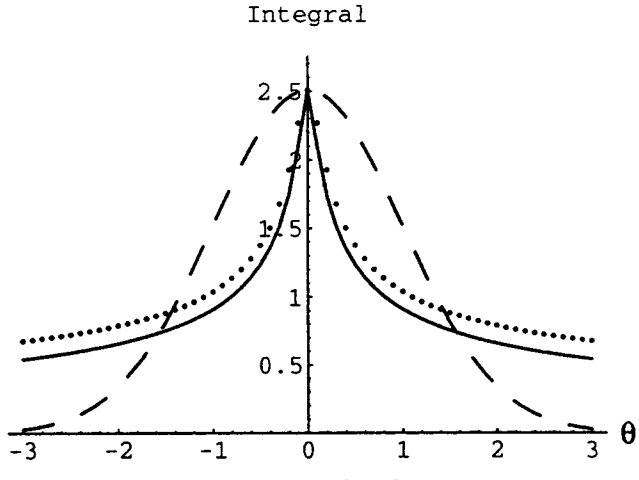

Figure 3. Exact integral on θ (——), Laplace approximation (· · · ·) and linear approximation (----) for noise on incidental parameters $\sigma = 2$.

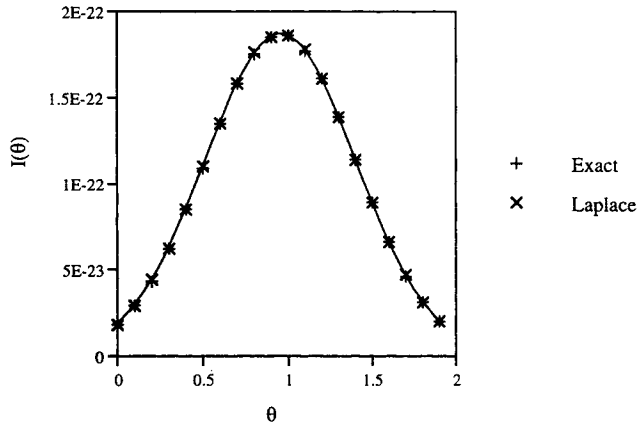

Figure 4. Exact density on θ and Laplace approximation.

$n = 10$ and $\sigma = 0.01$.

dental parameters) by expanding the number of parameters in the above problem. Consider the following measurements

$$\begin{aligned} \bar{x}_{1,i} &= x_{1,i} + \epsilon_x, \quad \epsilon_x \sim N(0,1) \\ \bar{x}_{2,i} &= x_{2,i} + \epsilon_x, \quad \epsilon_x \sim N(0,1) \\ \bar{u}_i &= u_i + \epsilon_u, \quad \epsilon_u \sim N(0,\sigma^2) \end{aligned} \tag{43}$$

where the states $x_{1,i}$ and $x_{2,i}$, the controls u_i, and the parameter θ are related through

$$x_{1,i} = \theta u_i^2 \tag{44}$$

$$x_{2,i} = \theta c_i^2 \tag{45}$$

$$i = 1, \cdots, n$$

and c_i are constants. This example was simulated by setting $\theta = 1$, $u_i = i/n$, $c_i = i/n$ and computing the corresponding $x_{1,i}$ and $x_{2,i}$. The data were simulated by adding randomly generated noise to the simulated values according to the model in

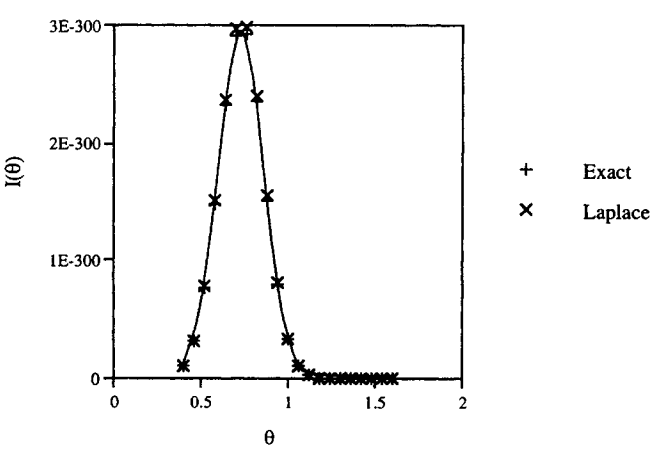

Figure 5. Exact density on θ and Laplace approximation.

$n = 150$ and $\sigma = 0.01$.

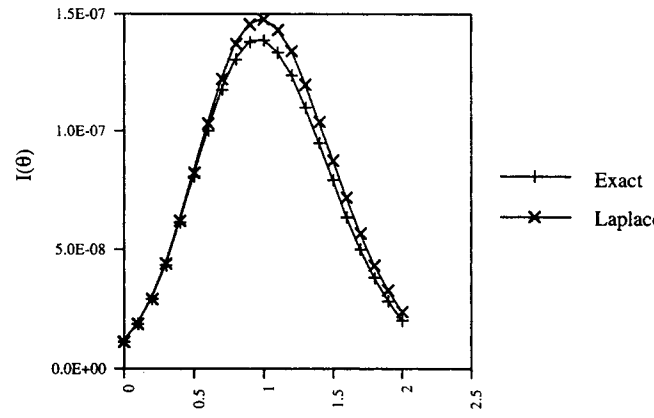

Figure 6. Exact density on θ and Laplace approximation.

$n = 10$ and $\sigma = 0.3$.

Eq. 43. By using Eqs. 44 and 45, we get the following joint posterior density in the parameters and the inputs

$$p(\theta, u_i, \cdots, u_n | \bar{x}, \bar{u}) \propto \prod_{i=1}^{n} \exp\left[-\frac{1}{2}(\theta u_i^2 - \bar{x}_{1i})^2 - \frac{1}{2}(\theta c_i^2 - \bar{x}_{2i})^2 - \frac{1}{2}\left(\frac{u_i - \bar{u}_i}{\sigma}\right)^2 \right] \tag{46}$$

and the marginal density in θ becomes

$$p(\theta | \bar{x}, \bar{u}) = \int_{-\infty}^{+\infty} p(\theta, u_1, \cdots, n) du_1 \cdots du_n. \tag{47}$$

Figures 4 and 5 compare the integral in Eq. 47 obtained by the numerical integration with the Laplace approximation for a small number of data sets ($n = 10$) and a large number of data sets ($n = 150$) for the case of a small measurement noise in the inputs ($\sigma = 0.01$). In Figures 6 and 7, we do the same comparison but for the case of noisier inputs ($\sigma = 0.3$).

For $\sigma = 0.01$, the approximation in Eq. 29 holds even for a large number of incidental parameters, but for $\sigma = 0.3$ the

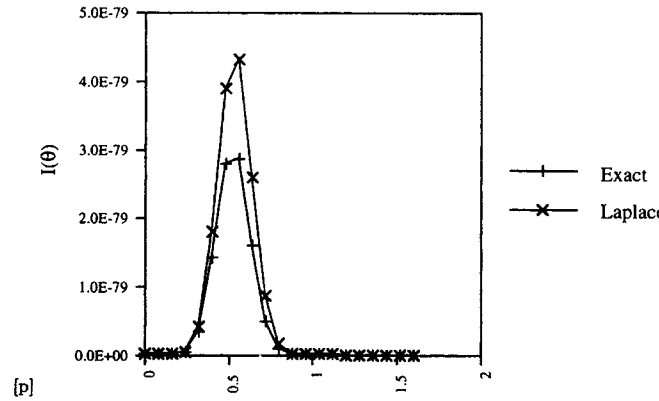

Figure 7. Exact density on θ and Laplace approximation.

$n = 150$ and $\sigma = 0.3$.

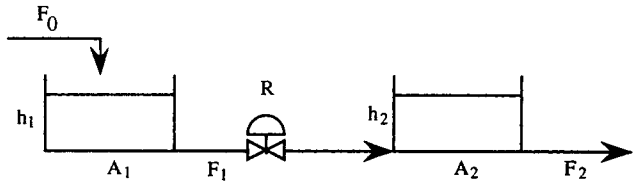

Figure 8. Two tanks connected by a valve.

inputs are noisy enough so that the approximation breaks down even for a few points and loses accuracy as the number of data sets increases. Thus, we conclude that the approximation in Eq. 29 is reliable for large numbers of inputs as long as the measurement noise level remains small.

Two connected tanks

In this example, which we treated previously (Albuquerque and Biegler, 1996a,b), we have two tanks connected by a valve. The measured variables are the flows F_0, F_1, F_2 and the levels of liquid h_1, h_2. The parameters are the inverses of the cross-sectional areas $1/A_1$, $1/A_2$ and the incidental parameters are the feed flows F_0. The arrangement is shown in Figure 8. The model for this process is described by the following DAE

$$A_1\dot{h}_1 = F_0 - F_1 \tag{48}$$

$$A_2\dot{h}_2 = F_1 - F_2 \tag{49}$$

$$h_1 = h_2 \tag{50}$$

$$F_2 = A_2\sqrt{2gh_2}. \tag{51}$$

The states and control variable were discretized with $n = 10$. To simplify the solution of these equations, Eq. 51 can be squared, since all variables involved (areas, flows, levels) are positive, becoming $-F_2^2/A_2^2 + 2gh_2 = 0$. Note that the only nonlinearity comes from this equation as its second derivative with respect to the state variable F_2 is equal to $-2/A_2^2$. The

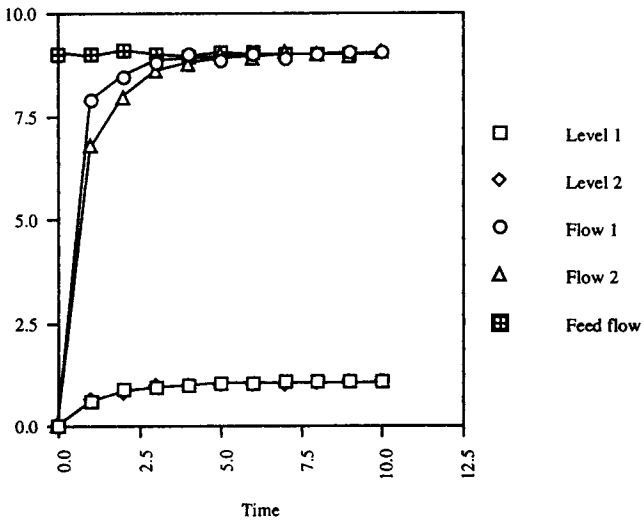

Figure 9. Data for the two tanks.

Levels in meters; flows in cubic per s.

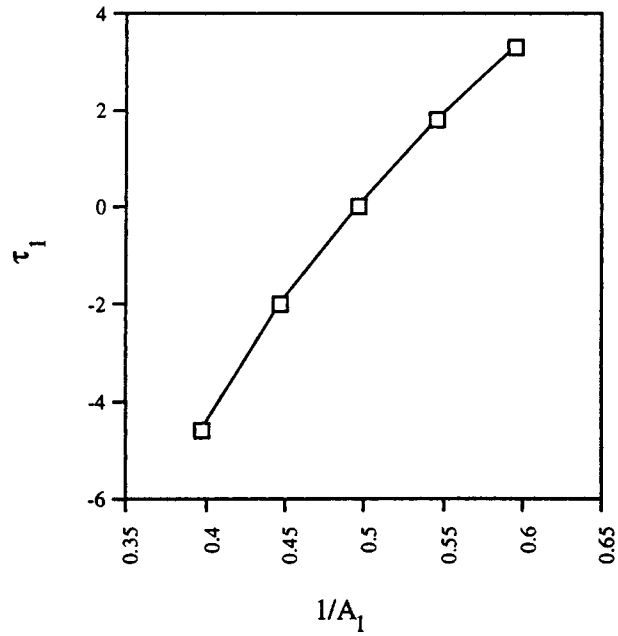

Figure 10. Profile τ plot for $1/A_1$.

data were simulated by solving the DAE system with $A_1 = A_2 = 2m^2$ and $F_0 = 9m^3/s$ and adding Gaussian noise, with a zero mean and a standard deviation of 0.05. These are displayed in Figure 9.

A least-squares objective function was used and the estimates for the parameters are 0.496 and 0.503. The nonlinearity in the model is very weak. The profile τ plots are represented in Figures 10 and 11, and the combined pairwise plot is shown in Figure 12. From these plots, marginal confidence intervals for $1/A_1$ and $1/A_2$ can be directly obtained by equating τ to $\pm z_{1-\alpha/2}$. For instance, intervals with a 99%, 95%, and 90% may be obtained by equating τ_1 and τ_2 to

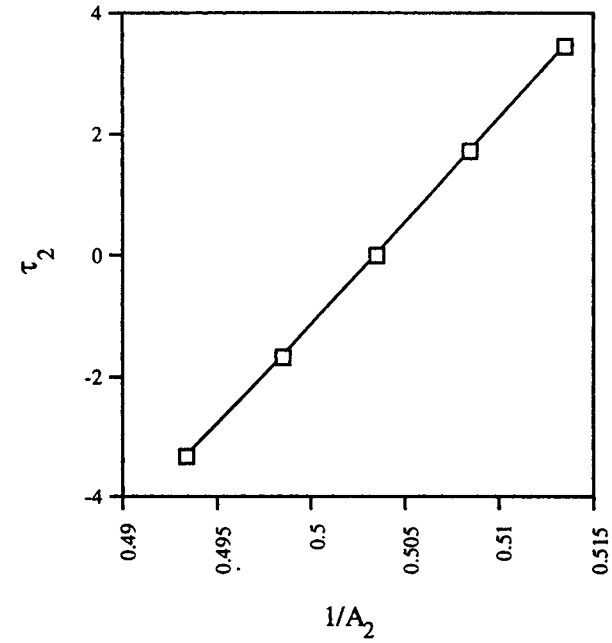

Figure 11. Profile τ plot for $1/A_2$.

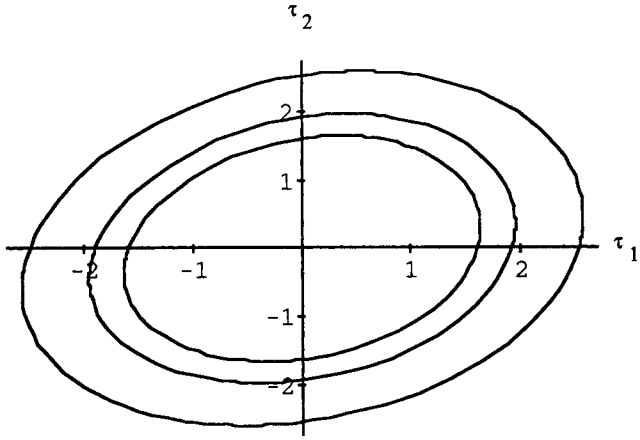

Figure 12. Likelihood contours for the tanks problem in the τ space.

Innermost: 90%; inner: 95%; outer: 99%.

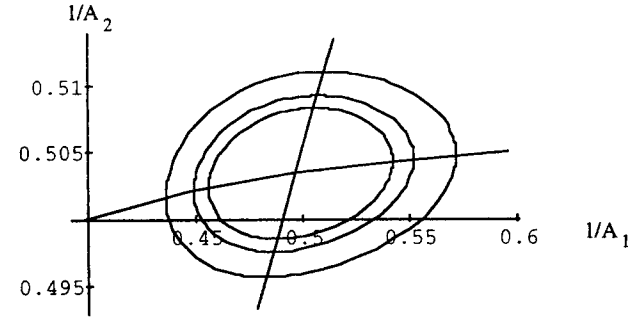

Figure 13. Likelihood contours for the tanks problem in the parameter space with the profile trace plot overlaid.

Innermost: 90%; inner: 95%; outer: 99%.

± 2.57, ± 1.96 and ± 1.64, and reading the corresponding values of $1/A_1$ and $1/A_2$ off the horizontal scale in Figures 10 and 11. For $1/A_1$ we get [0.436, 0.571], [0.449, 0.551] and [0.456, 0.542]. For $1/A_2$, we get [0.496, 0.511], [0.498, 0.509] and [0.499, 0.508]. In this problem, the correction for integrating the control variables was very small. Finally, Figure 13 shows the interpolated contours in the parameter space. Each regression took 7.8 s of CPU time in a MicroVAX 3200 workstation. Computing the second derivatives and their determinant took 5.5 s. Total running time was 146 s.

Reactor problem

In this problem, treated previously in Sistu et al. (1993) and Albuquerque and Biegler (1995), we consider a nonisothermal CSTR with an exothermal irreversible reaction. The equations that model the process in dimensionless quantities consist of the following system of ordinary differential equations (ODE)

$$\frac{dx_1}{\partial t} = -\Phi x_1 \exp\left(\frac{x_2}{1+x_2/\lambda}\right) + q(x_{1f} - x_1) \tag{52}$$

$$\frac{dx_2}{dt} = \beta\Phi x_1 \exp\left(\frac{x_2}{1+x_2/\lambda}\right) - (q+\delta)x_2 + \delta u + q x_{2f}. \tag{53}$$

Here x_1, x_2 are the dimensionless concentration and temperature of the exit stream, and they are state variables. The feed conditions correspond to x_{1f} and x_{2f}, and these are assumed to be constant and error free. The control variable u is the dimensionless temperature of the cooling jacket, t is the dimensionless time variable, Φ is the dimensionless Arrhenius factor, β is the dimensionless reaction enthalpy, δ is the dimensionless heat-transfer coefficient of the cooling jacket, λ is the activation energy, and q is the input to output flow ratio. These last five quantities are time independent and are known, except for Φ and λ (the kinetic parameters). The system was simulated with (Φ, β, δ, λ, q) set to (0.072, 8, 0.3, 20, 1.0), and x_{1f} and x_{2f} set to 1.0 and zero, respectively, and with u set to a flat zero profile. White noise with a standard-deviation of 0.1 was added to x_1, x_2, and u.

The data, together with the true values and the fitted model, are taken from Albuquerque and Biegler (1995). The states and control variable were discretized with $n = 100$, and the problem was solved using a least-squares objective function and the fitted values for Φ and λ were 0.0718 and 20.144. The profile trace plot for this problem is shown in Figure 16, and the profile τ plots are shown in Figures 14 and 15. Marginal confidence intervals of 99%, 95% and 90% for Φ are given as [0.0678, 0.0758], [0.0687, 0.0749], and [0.0692, 0.0744], respectively. For λ we get [16.24, 26.21], [16.99, 24.51] and [17.40, 23.91]. A plot of the confidence contours in the parameter space overlaid with the profile trace plot is given in Figure 17. In this problem, each regression took 0.13 s in a MicroVAX 3200 workstation. Computing second derivatives and their determinant took 1 s. The total CPU time was 26.4 s. This problem shows a strong degree of nonlinearity which would not have been captured using linear approximations.

We can check Laplace's method in this example by comparing the profiles obtained in the absence of incidental parameters with those obtained in the presence of a very small measurement error in the incidental parameters. In both cases, the measurements for the state variables x were simulated by adding normally distributed noise with a standard-deviation of 0.1 to the true values. In the last case, the measurements for the control variable u were generated using normally distributed noise with a standard deviation of 0.0001.

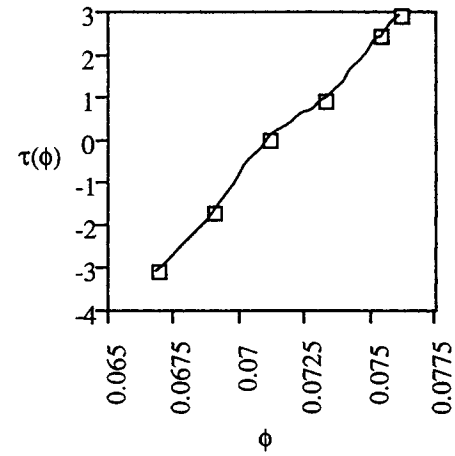

Figure 14. Profile τ plot for parameter ϕ.

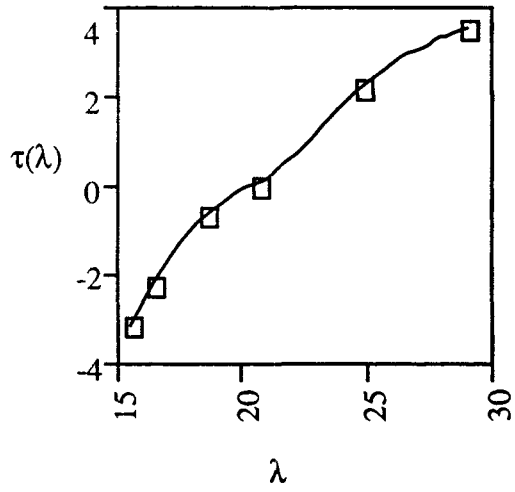

Figure 15. Profile τ plot for parameter γ.

The results are so close that the graphs overlap. More detailed information on this case can be found in Albuquerque (1996). Finally, if we compare the combined profile trace plots for the case where there are no incidental parameters, the correlation between the parameters is so strong that both curves overlap. Here it is evident that the Laplace correction is negligible for the case when the noise in the incidental parameters is very small.

Conclusions

EVM problems arise in many steady-state and dynamic parameter estimation problems. These problems are especially important for on-line applications in process control where measurement errors in the process outputs *and inputs* cannot be ignored. Efficient algorithms have recently been developed to solve nonlinear EVM problems (Albuquerque and Biegler, 1995, 1996a,b) but without the capability for nonlinear inference of these parameters. Also, questions of dependence among parameters and parametric sensitivity remain unanswered, and these considerations may even invalidate the results of the point estimation.

On the other hand, exact methods for nonlinear inference are expensive for process models even if only a handful of

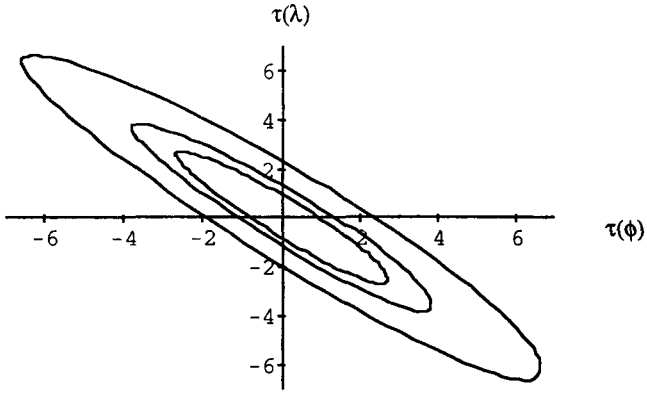

Figure 16. Likelihood contours for the reactor problem in the τ space.

Innermost: 90%; inner: 95%; outer: 99%.

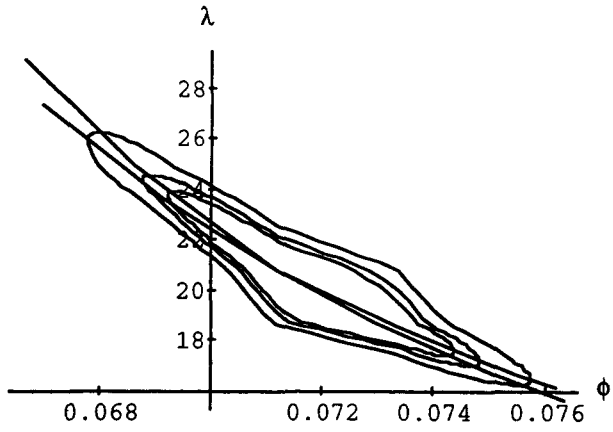

Figure 17. Likelihood contours for the reactor problem in the parameter space with the profile trace plot overlaid.

Innermost: 90%; inner: 95%; outer: 99%.

parameters are present. Instead, for this case, an efficient profiling approach (Bates and Watts, 1988) leads to a reasonably accurate approximation of confidence regions. However, when the number of parameters becomes large, as in the EVM problem, even this approach is prohibitively expensive.

In this study, we have proposed a new approach for drawing inferences from EVM models in dynamic systems, and we derived and demonstrated a computationally feasible method. This approach combines profiling and Laplace's method, both of which may be justified by asymptotic arguments. We turn to these approximations because of the computational burden inherent in dynamic systems; each evaluation of the likelihood function is costly because a DAE system must be solved. As a result, Monte Carlo or quadrature methods (Genz and Kass, 1997 and the references therein) cannot be used in this setting.

Although Laplace's method can break down as the number of incidental parameters u_i increases (Shun and McCullagh, 1994), our examples demonstrate that the technique remains effective when the noise in the incidental parameters is small, even for a large number of control variables. In practice, the control and input variables are usually selected so as to be measured fairly accurately so that the process may be controlled in a reliable way. Thus, we believe the Laplace approximation is a reasonable method to use in these problems.

Finally, this study indicates that estimation of and nonlinear inference about process parameters can be obtained fairly inexpensively by applying profiling and Laplace's approximation. Its performance on two nonlinear process systems indicates that this approach can be coupled with efficient on-line state and parameter estimation algorithms with reasonable computational requirements. As a result, this approach leads to an efficient and effective analysis tool for process modeling, data reconciliation and on-line optimization.

Literature Cited

Albuquerque, J. S., PhD Thesis, Chemical Engineering Dept., Carnegie Mellon University, Pittsburgh, PA (1996).

Albuquerque, J. S., and L. T. Biegler, "Decomposition Algorithms for On-Line Estimation with Nonlinear Models," *Comput. Chem. Eng.*, **19**(10), 1031 (1995).

Albuquerque, J. S., and L. T. Biegler, "Decomposition Algorithms for On-Line Estimation with Nonlinear DAE Models," *Comput. Chem. Eng.*, **21**(3), 283 (1997).

Albuquerque, J. S., and L. T. Biegler, "Data Reconciliation and Gross-Error Detection for Dynamic Systems," *AIChE J.*, **42**(10), 2841 (1996).

Basu, A., and K. K. Paliwal, "Robust M-Estimates and Generalized M-Estimates for Autoregressive Parameter Estimation," *TENCON 89, Fourth IEEE Reg. 10 Int. Conf.*, Bombay, India (1989).

Bates, D. M., and D. G. Watts, *Nonlinear Regression Analysis and its Applications*, Wiley (1988).

Betts, J. T., "Experience with a Sparse Nonlinear Programming Algorithm," IMA Workshop on Large Scale Optimization, University of Minnesota (1995).

Crowe, C. M., "Observability and Redundancy of Process Data for Steady State Reconciliation," *Chem. Eng. Sci.*, **44**(12), 2909 (1989).

Fariss, R. H., and V. H. Law, "An Efficient Computational Technique for Generalized Application of Maximum Likelihood to Improve Correlation of Experimental Data," *Comput. Chem. Eng.*, **3**, 95 (1979).

Forsythe, G. E., *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, NJ (1977).

Genz, A., and R. K. Kass, "Subregion-Adaptive Integration of Functions having a Dominant Peak," *J. Computational and Graphical Stat*, (in press, 1997).

Johnston, L. P. M., and A. M. Kramer, "Maximum Likelihood Data Rectification: Steady State Systems," *AIChE J.*, **41**(11), 2145 (1995).

Kass, R. K., L. Tierney, and J. B. Kadane, "Laplace's Method in Bayesian Analysis," *Contemp. Mathemat.*, **115**, 89 (1991).

Kretsovalis, A., and R. S. H. Mah, "Observability and Redundancy Classification in Generalized Process Networks—I. Theorems," *Comput. Chem. Eng.*, **12**, 671 (1988).

Liebman, M. J., T. F. Edgar, and L. S. Lasdon, "Efficient Data Reconciliation and Estimation for Dynamic Processes using Nonlinear Programming Techniques," *Comp. Chem. Eng.*, **16**(10/11), 963 (1992).

Murray, J. D., *Asymptotic Analysis*, Clarendon Press (1974).

Narasimhan, S., and R. S. H. Mah, "Generalized Likelihood Ratios for Gross Error Identification in Dynamic Processes," *AIChE J.*, **34**(8), 1321 (1988).

Schlöder, J. P., M. W. Ziesse, H. G. Bock, and J. V. Gallitzendörfer, "Parameter Estimation in Multispecies Transport Reaction Systems using Parallel Algorithms," Preprint 95-26, Interdisciplinäres Zentrum Fü Wissenschaftliches Rechnen der Universität Heidelberg, Germany (1995).

Seber, G. A., and C. J. Wild, *Nonlinear Regression*, Wiley, New York (1989).

Shun, Z., and P. McCullagh, "Laplace Approximation of High-Dimensional Integrals," Technical Report no. 389, Dept. of Statistics, The University of Chicago, Chicago (1994).

Sistu, P. B., R. S. Gopinath, and B. W. Bequette, "Computation Issues in Nonlinear Predictive Control," *Comp. Chem. Eng.*, **17**(4), 361 (1993).

Steinbach, M., "Fast Recursive SQP Methods for Large-Scale Optimal Control Problems," PhD Thesis, Interdisciplinäres Zentrum Fü Wissenschaftliches Rechnen der Universität Heidelberg, Germany (1995).

Swartz, C. L. E., "Data Reconciliation for Generalized Flowsheet Applications," paper presented at the American Chemical Society National Meeting, Dallas, TX (Apr. 1989).

Sweeting, T. J., "A Framework for Bayesian and Likelihood Approximations in Statistics," *Biometrika*, **82**, 1 (1995).

Tamhane, A. C., C. Kao, and R. S. H. Mah, "Gross Error Detection in Serially Correlated Process Data. 2. Dynamic Systems," *Ind. Eng. Chem. Res.*, **31**, 254 (1992).

Tierney, L., and J. B. Kadane, "Accurate Approximations for Posterior Moments and Marginal Densities," *J. Amer. Statist. Assoc.*, **81**, 82 (1986).

Tjoa, I. B., and L. T. Biegler, "Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems," *Comput. Chem. Eng.*, **15**(10), 679 (1991).

Wolfram, S., *Mathematica. A System for Doing Mathematics by Computer*, Addison-Wesley, Reading, MA (1988).

Manuscript received July 8, 1996, and revision received Nov. 18, 1996.