Article

Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation

Eric L. Haseltine, and James B. Rawlings

Ind. Eng. Chem. Res., **2005**, 44 (8), 2451-2460 • DOI: 10.1021/ie034308l

Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation

Eric L. Haseltine and James B. Rawlings*

Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, Wisconsin 53706-1607

State estimators for physical processes often must address different challenges, including nonlinear dynamics, states subject to hard constraints (e.g., nonnegative concentrations), and local optima. In this article, we compare the performance of two such estimators: the extended Kalman filter (EKF) and moving-horizon estimation (MHE). We outline conditions that lead to the formation of multiple optima in the estimator for systems tending to a steady state and propose tests that determine when these conditions hold for chemical reaction networks. Several simulation examples demonstrate estimation failure in the EKF, even in the absence of plant−model mismatch. We then examine the role that constraints play in determining the performance on these examples of MHE employing local optimization and a "smoothing" update for the arrival cost. This implementation of MHE represents a feasible, on-line alternative to the EKF for industrial practitioners. In each example, the two estimators are given exactly the same information, namely, tuning parameters, model, and measurements; yet MHE consistently provides improved state estimation and greater robustness to both poor guesses of the initial state and tuning parameters in comparison to the EKF. The only price of this improvement is the greater computational expense required to solve the MHE optimization.

1. Introduction

It is well established that the Kalman filter is the optimal state estimator for unconstrained, linear systems subject to normally distributed state and measurement noise. Many physical systems, however, exhibit nonlinear dynamics and have states subject to hard constraints, such as nonnegative concentrations or pressures. In these cases, Kalman filtering is no longer directly applicable. As a result, many different types of nonlinear state estimators have been proposed; Soroush[1] provides a review of many of these methods. We focus our attention on techniques that formulate state estimation in a probabilistic setting, that is, both the model and the measurement are potentially subject to random disturbances. Such techniques include the extended Kalman filter, moving-horizon estimation, Bayesian estimation, and Gaussian sum approximations. In this probabilistic setting, state estimators attempt to reconstruct the a posteriori distribution $p(x_T|y_0,...,y_T)$, which is the probability that the state of the system is x_T given measurements y_0, ..., y_T. The question arises, then, as to which *point* estimate should be used for the state estimate. Two obvious choices for the point estimate are the mean and the mode of the a posteriori distribution. For nonsymmetric distributions, Figure 1a demonstrates that these estimates are generally different. Additionally, if this distribution is multimodal, as is Figure 1b, then the mean might place the state estimate in a region of low probability. Clearly, the mode is a more desirable estimate in such cases.

For nonlinear systems, the a posteriori distribution is generally nonsymmetric and potentially multimodal. In this paper, we outline conditions that lead to the formation of multiple modes in the a posteriori distribution for systems tending to a steady state and construct examples that generate multiple modes. To the best of our knowledge, only Alspach and Sorenson (and references contained therein);[2] Gordon, Salmond, and Smith;[3] and Chaves and Sontag[4] have proposed examples in which multiple modes arise in the a posteriori distribution, but these contributions do not examine conditions leading to their formation. Gaussian sum approximations[2] offer one method for addressing the formation of multiple modes in the a posteriori distribution for unconstrained systems. Current Bayesian estimation methods[3,5−7] offer another means for addressing multiple modes, but these methods propose estimation of the mean rather than the mode. Gordon, Salmond, and Smith[3] suggest using density estimation techniques to estimate the mode of the a posteriori distribution, but Silverman[8] demonstrates via a numerical example that the number of samples required to reconstruct a point estimate within a given relative error increases exponentially with the dimensionality of the state. In this paper, we examine the estimation properties of both the extended Kalman filter and moving-horizon estimation through simulation. The extended Kalman filter assumes that the a posteriori distribution is normally distributed (unimodal); hence, the mean and the mode of the distribution are equivalent. Moving-horizon estimation seeks to reconstruct the mode of the a posteriori distribution via constrained optimization, but current implementations employ local optimizations that offer no means of distinguishing between multiple modes of this distribution. The simulation examples thus provide a means of benchmarking these current industrially implementable technologies.

In this paper, we first formulate the estimation problem of interest. Next, we briefly review pertinent extended Kalman filtering and moving-horizon estimation literature. Then, we present several motivating chemical engineering examples in which the accurate

* To whom correspondence should be addressed. E-mail: jbraw@bevo.che.wisc.edu. Fax: (608) 265-8794.

10.1021/ie034308l CCC: $30.25
Published on Web 06/29/2004

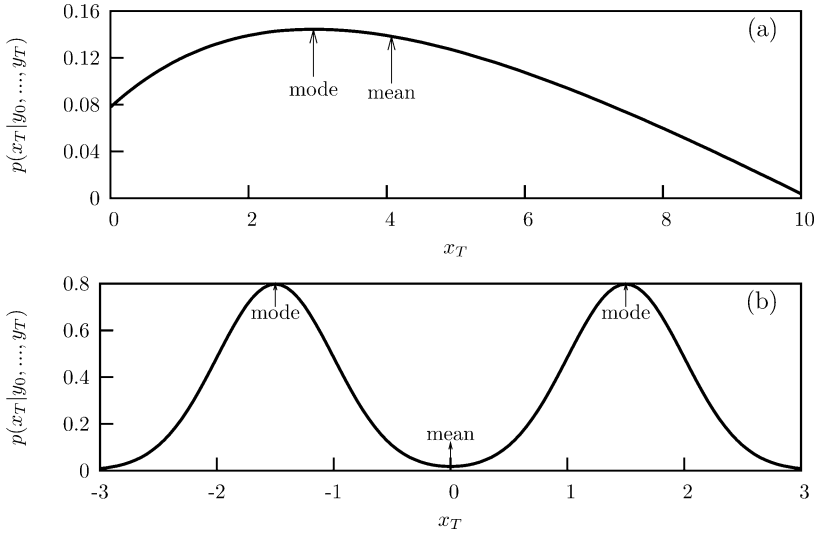

Figure 1. Comparison of potential point estimates (mean and mode) for (a) unimodal and (b) bimodal a posteriori distributions.

incorporation of both state constraints and the nonlinear model are paramount for obtaining accurate estimates.

2. Formulation and Solution of the Estimation Problem

In chemical engineering systems, most processes consist of continuous processes with discrete measurements. Therefore, for this work, we choose the discrete stochastic system model

$$x_{k+1} = F(x_k,u_k) + G(x_k,u_k)w_k \tag{1a}$$

$$y_k = h(x_k) + v_k \tag{1b}$$

in which x_k is the state of the system at time t_k; u_k is the system input at time t_k {assumes a zeroth-order hold over the interval $[t_k, t_{k+1})$}; w_k is an $\mathcal{N}(0,Q_k)$ noise [$\mathcal{N}(m,P)$ denotes a normal distribution with mean m and covariance P]; $F(x_k,u_k)$ is the solution to a first-principles, differential equation model; $G(x_k,u_k)$ is a full column rank matrix (this condition is required for uniqueness of the a posteriori distribution defined in section 2.2); y_k is the system measurement at time t_k; h_k is a (possibly) nonlinear function of x_k at time t_k; and v_k is an $\mathcal{N}(0,R_k)$ noise.

We believe that, by appropriately choosing both a first-principles model and a noise structure, we can identify both the model parameters (or a reduced set of these parameters) and the state and measurement noise covariance structures. Such identification will proceed as follows: (1) Assuming a noise structure, identify the model parameters. (2) Assuming a model, model parameters, and a noise structure, identify the covariance structures.

Here, we propose performing replicate experiments and measurements to estimate moments of the desired quantity (in general, the mean of the state or covariance structure) and then fitting the model parameters by comparing the estimated moments to those reconstructed from Monte Carlo simulation of eq 1. This identification procedure is an area of current research beyond the scope of this paper, but we maintain that such a procedure will yield a rough, potentially biased, yet useful stochastic model from the system measurements.

As discussed in the Introduction, state estimators given multimodal a posteriori distributions should solve the problem

$$x_T^+ = \arg\max_{x_T} p(x_T|y_0,...y_T) \tag{2}$$

Here, we assume that the input sequence u_0, ..., u_T is known exactly. Equation 2 is referred to as the maximum a posteriori estimate. In the special case that the system is not constrained and that, in eq 1, (1) $F(x_k,u_k)$ is linear with respect to x_k, (2) $h(x_k)$ is linear with respect to x_k, and (3) $G(x_k,u_k)$ is a constant matrix, the maximum a posteriori estimator is the Kalman filter, whose well-known recursive form is conducive for online implementation. For the more general formulation given by eq 1, online solution of the exact maximum a posteriori estimate is impractical, and approximations are used to obtain state estimates in real time. We consider two of these approximations, the extended Kalman filter and moving-horizon estimation, next.

2.1. Extended Kalman Filtering. The extended Kalman filter, or EKF, is one approximation for calculating eq 2. The EKF linearizes the nonlinear system and then applies the Kalman filter (the optimal, unconstrained, linear state estimator) to obtain the state estimates. The tacit approximation here is that the process statistics are multivariate normal distributions. This method has garnered the most interest because of its relative simplicity and demonstrated efficacy in handling nonlinear systems. Examples of implementations include estimation for the production of silicon/germanium alloy films,[9] polymerization reactions,[10] and fermentation processes.[11] However, the EKF is at best an ad hoc solution to a difficult problem, and hence, there exist many barriers to the practical implementation of EKFs (see, for example, Wilson, Agarwal, and Rippin[12]). Some of these problems include the inability to accurately incorporate physical state constraints and poor use of the nonlinear model.

Until recently, few properties regarding the stability and convergence of the EKF have been proven. Recent publications present bounded estimation error and exponential convergence arguments for the continuous and discrete EKF forms given detectability, small initial estimation error, small noise terms, and perfect correspondence between the plant and the model.[13−15]

However, depending on the system, the bounds on initial estimation error and noise terms might be unreasonably small. Also, initial estimation error might result in bounded estimate error but not exponential convergence, as illustrated by Chaves and Sontag.[4] We use the discrete-time EKF (see Stengel[16] for a summary of the algorithm) to generate the results in this paper.

2.2. Moving-Horizon Estimation. One alternative to solving the maximum a posteriori estimate is to maximize a joint probability for a trajectory of state values, i.e.

$$\{x_0^*, ..., x_T\} = \arg \max_{x_0,...,x_T} p(x_0,...,x_T|y_0,...,y_T) \quad (3)$$

Equation 3 is the full-information estimate. The computational burden of calculating this estimate increases as more measurements come online. To bound this burden, one can fix the estimation horizon

$$\{x_{T-N+1}, ..., x_0\} = \arg \max_{x_{T-N+1},...,x_T} p(x_{T-N+1},...,x_T|y_0,...,y_T) \quad (4)$$

Moving-horizon estimation, or MHE, corresponds probabilistically to eq 4 and is equivalent numerically to a constrained, nonlinear optimization problem.[17,18] We note that the restrictive assumptions of normally distributed noises and the model given by eq 1 are not required by MHE. If the matrix G in eq 1 is not a function of the state x_k, then these assumptions merely lead to a convenient least-squares optimization as demonstrated by Jazwinski.[19]

From a theoretical perspective, Tyler and Morari examined the feasibility of constrained MHE for linear, state-space models.[20] Rao et al. showed that constrained MHE is an asymptotically stable observer in a nonlinear deterministic modeling framework.[21,22] These works also provide a nice overview of current MHE research. Furthermore, recent advances in numerical computation have allowed real-time implementation of MHE strategies for the local optimization of the MHE problem.[23,24] A method for incorporating the effect of past data outside the current estimation horizon (also known as the arrival cost), though, remains an open issue of MHE.

Rao, Rawlings, and Lee[25] explored estimating this cost for constrained linear systems with the corresponding cost for an unconstrained linear system. More specifically, they examinedthe following two schemes: (1) a "filtering" scheme that penalizes deviations of the initial estimate in the horizon from an a priori estimate, and (2) a "smoothing" scheme that penalizes deviations of the trajectory of states in the estimation horizon from an a priori estimate.

For unconstrained, linear systems, the MHE optimization collapses to the Kalman filter for both of these schemes. Rao[21] further considered several optimal and suboptimal approaches for estimating the arrival cost via a series of optimizations. These approaches stem from the property that, in a deterministic setting (no state or measurement noise), MHE is an asymptotically stable observer as long as the arrival cost is underbounded. One simple way of estimating the arrival cost, therefore, is to implement a uniform prior. Computationally, a uniform prior corresponds to not penalizing deviations of the initial state from the a priori estimate.

For nonlinear systems, Tenny and Rawlings[24] estimated the arrival cost by approximating the constrained, nonlinear system as an unconstrained, linear time-varying system and applying the corresponding filtering and smoothing schemes. They concluded that the smoothing scheme is superior to the filtering scheme because the filtering scheme induces oscillations in the state estimates through the unnecessary propagation of initial error. Here, the tacit assumption is that the probability distribution around the optimal estimate is a multivariate normal. The problem with this assumption is that nonlinear systems can exhibit multiple peaks (i.e., local optima) in this probability distribution. Haseltine and Rawlings[26] demonstrated that approximating the arrival cost with the smoothing scheme in the presence of multiple local optima can skew all future estimates. They conjectured that, if global optimization is implementable in real time, approximating the arrival cost with a uniform prior and making the estimation horizon reasonably long is preferable to using an approximate multivariate normal arrival cost because of the latter's biasing effect on the state estimates.

We now seek to demonstrate by simulation examples that MHE is a useful and practical tool for state estimation of chemical process systems. We examine the performance of MHE with local optimization and an arrival cost approximated with a smoothing update. For further details regarding this MHE scheme, we refer the interested reader to Tenny and Rawlings[24] and note that this code is freely available as part of the NMPC toolbox (http://www.che.wisc.edu/~tenny/nmpc/). Currently, this particular MHE configuration represents a computationally feasible implementation for an industrial setting.

3. Examples of EKF Failure

In this section, we outline the conditions that generate EKF failure in two classes of chemical reactors. We then present several examples that demonstrate failure of the EKF as an estimator.

If there is no plant−model mismatch, measurement noise, or state noise, one definition of estimator failure is

$$\lim_{k\to\infty} |\hat{x}_{k|k} - x_k| > \epsilon \quad (5)$$

for some $\epsilon > 0$ ($|x|$ is a norm of x). That is, the estimator is unable to reconstruct the true state no matter how many measurements it processes. For stable systems, i.e., those systems tending to a steady state, we expect that

$$\hat{x}_{k|k} = \hat{x}_{k-1|k-1} \quad (6)$$

in the same limit as eq 5. We now examine the discrete EKF given such conditions. The following equations govern the propagation and update steps:[16]

$$\hat{x}_{k|k-1} = F(\hat{x}_{k-1|k-1}, u_{k-1}, w_{k-1}) \quad (7a)$$

$$P_{k|k-1} = A_{k-1} P_{k-1|k-1} A_{k-1}^{\mathrm{T}} + G_{k-1} Q_{k-1} G_{k-1}^{\mathrm{T}} \quad (7b)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k[y_k - h(\hat{x}_{k|k-1})] \quad (7c)$$

$$P_{k|k} = P_{k|k-1} - L_k C_k P_{k|k-1} \quad (7d)$$

$$L_k = P_{k|k-1} C_k^{\mathrm{T}} [C_k P_{k|k-1} C_k^{\mathrm{T}} + R_k]^{-1} \quad (7e)$$

in which

$$A_k = \frac{\partial F(x_k,u_k,w_k)}{\partial x_k^{\mathrm{T}}} \tag{8a}$$

$$G_k = \frac{\partial F(x_k,u_k,w_k)}{\partial w_k^{\mathrm{T}}} \tag{8b}$$

$$C_k = \frac{\partial h(x_k)}{\partial x_k^{\mathrm{T}}} \tag{8c}$$

At steady state, the following equalities hold:

$$\hat{x}_{k|k} = \hat{x}_{k-1|k-1} \tag{9a}$$

$$P_{k|k} = P_{k-1|k-1} \tag{9b}$$

Combining expressions 7 and 9 yields

$$0 = F(\hat{x}_{k-1|k-1},u_{k-1}) - \hat{x}_{k|k-1} \tag{10a}$$

$$0 = A_{k-1}P_{k-1|k-1}A_{k-1}^{\mathrm{T}} + G_{k-1}Q_{k-1}G_{k-1}^{\mathrm{T}} - P_{k|k-1} \tag{10b}$$

$$0 = \hat{x}_{k|k-1} + L_k[y_k - h(\hat{x}_{k|k-1})] - \hat{x}_{k-1|k-1} \tag{10c}$$

$$0 = P_{k|k-1} - L_kC_kP_{k|k-1} - P_{k-1|k-1} \tag{10d}$$

$$L_k = P_{k|k-1}C_k^{\mathrm{T}}[C_kP_{k|k-1}C_k^{\mathrm{T}} + R_k]^{-1} \tag{10e}$$

If both eqs 5 and 10 hold, then the EKF has failed as an estimator.

One solution to eq 10 results when multiple steady states satisfy the steady-state measurement. This phenomenon corresponds to the case that

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} = \hat{x}_{k-1|k-1} \tag{11}$$

$$y_k = h(\hat{x}_{k|k-1}) \tag{12}$$

$$\hat{x}_{k|k} \neq x_k \tag{13}$$

We would expect the EKF to fail when (1) the system model and measurement are such that multiple states satisfy the steady-state measurement and (2) the estimator is given a poor initial guess of the state. Condition 2 does not imply that the system is unobservable; rather, this condition states that the state cannot be uniquely determined from solely the *steady-state measurement*. For such a case to be observable, the process dynamics must make the system observable. Condition 2 implies that the poor initial guess skews the estimates ($\hat{x}_{k|k}$'s) toward a region of attraction not corresponding to the actual state (x_k's).

3.1. Chemical Reaction Systems. For well-mixed systems consisting of reaction networks, the nonlinearity of the system must be present at steady state so that multiple steady states can satisfy the steady-state measurement. Consequently, we must analyze the structure of the stoichiometric matrix in combination with the number (and type) of measurements to determine whether multiple steady states can satisfy the steady-state measurement. Define the following quantities:

ν, the stoichiometric matrix of size $r \times s$, in which r is the number of reactions and s is the number of species

ρ, the rank of ν ($\rho = r$ if there are no linearly dependent reactions)

η, the nullity of ν

n, the number of measurements

n_{m}, the number of measurements that can be written as a linear combination of states [e.g., $y = x_1 + x_2$ and $(x_1 + x_2)y = x_1$]

For batch reactors, conservation laws yield a model of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}(xV_{\mathrm{R}}) = \nu^{\mathrm{T}}r(x)V_{\mathrm{R}} \tag{14}$$

in which x is an s-vector containing the concentration of each species in the reactor, V_{R} is the volume of the reactor, and $r(x)$ is an r-vector containing the reaction rates.

For this system, ρ specifies the number of independent equations at equilibrium. In general, we will require that (1) all reactions be reversible and (2) the following inequalities hold

number of "linear" equations <
number of estimated species ≤
number of independent equations

i.e.

$$n_{\mathrm{m}} + \eta < s \leq n + \rho$$

Note that the batch reactor preserves the nonlinearity of the reaction rates in the steady-state calculation. Also, the combination of batch steady-state equations and measurements might or might not be an overspecified problem.

For continuously stirred tank reactors (CSTRs), conservation laws yield a model of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}(xV_{\mathrm{R}}) = Q_{\mathrm{f}}c_{\mathrm{f}} - Q_{\mathrm{o}}x + \nu^{\mathrm{T}}r(x)V_{\mathrm{R}} \tag{15}$$

where x is an s-vector containing the concentration of each species in the reactor, V_{R} is the volume of the reactor, Q_{f} is the volumetric flow rate into the reactor, c_{f} is an s-vector containing the inlet concentration of each species, Q_{o} is the effluent volumetric flow rate, and $r(x)$ is an r-vector containing the reaction rates.

Here, η specifies the number of linear algebraic relationships among the s species at equilibrium because the null space represents linear combinations of the material balances that eliminate nonlinear reaction rates. We will require i.e.

number of linear equations <
number of estimated species

$$n_{\mathrm{m}} + \eta < s \tag{16}$$

If expression 16 is an equality instead of an inequality, then determination of the steady state is generally a well-defined, linear problem with a unique solution. Note that the left-hand side of expression 16 is actually an upper bound because we could potentially choose a measurement contained within the span of the null space (a linear combination of the null vectors). However, such measurements would be invariant and hence would give no dynamic information. Also, expression 16 does not imply that multiple steady states can satisfy the steady-state measurement; rather, having multiple

steady states that can satisfy the steady-state measurement implies that expression 16 holds. EKF failure for CSTRs modeled by eq 15 must be confirmed by verifying that eqs 10 hold. This requirement differs from the batch case because, in general, the CSTR design equation (eq 15) yields a sufficient number of equations to calculate all possible steady states, whereas the batch design equation (eq 14) does not.

We now examine several examples that illustrate these points.

3.2. Example 1. Consider the gas-phase, reversible reactions

$$\mathrm{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{B} + \mathrm{C} \tag{17a}$$

$$2\mathrm{B} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \mathrm{B} + \mathrm{C} \tag{17b}$$

$$k = [0.5 \quad 0.05 \quad 0.2 \quad 0.01] \tag{17c}$$

with stoichiometric matrix

$$\nu = \begin{bmatrix} -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix} \tag{18}$$

and reaction rates

$$r = \begin{bmatrix} k_1 c_\mathrm{A} - k_2 c_\mathrm{B} c_\mathrm{C} \\ k_3 {c_\mathrm{B}}^2 - k_4 c_\mathrm{C} \end{bmatrix} \tag{19}$$

We define the state and measurements to be

$$x = [c_\mathrm{A} \quad c_\mathrm{B} \quad c_\mathrm{C}]^\mathrm{T} \tag{20a}$$

$$y = [RT \quad RT \quad RT]x \tag{20b}$$

where c_j denotes the concentration of species j in moles per liter, R is the ideal gas constant, and T is the reactor temperature in Kelvin. For the simulations, $RT = 32.84$ mol atm/L. We assume that the ideal gas law holds. We consider state estimation for both a batch reactor and a CSTR.

3.2.1. Batch Reactor. From first principles, the model for a well-mixed, constant-volume, isothermal batch reactor is

$$3\dot{x} = f(x) = \nu^\mathrm{T} r \tag{21}$$

$$x_0 = [0.5 \quad 0.05 \quad 0]^\mathrm{T} \tag{22}$$

We consider state estimation with the following parameters:

$$\Delta t = t_{k+1} - t_k = 0.25 \tag{23a}$$

$$\Pi_0 = \mathrm{diag}(0.5^2, 0.5^2, 0.5^2) \tag{23b}$$

$$G_k = \mathrm{diag}(1, 1, 1) \tag{23c}$$

$$Q_k = \mathrm{diag}(0.001^2, 0.001^2, 0.001^2) \tag{23d}$$

$$R_k = 0.25^2 \tag{23e}$$

$$\bar{x}_0 = [0 \quad 0 \quad 4]^\mathrm{T} \tag{23f}$$

The distribution for the initial state is $\mathcal{N}(\bar{x}_0,\Pi_0)$. Note that the initial guess, $\bar{x}_0$, is poor. The actual plant experiences $\mathcal{N}(0,Q_k)$ noise in the state and $\mathcal{N}(0,R_k)$ noise in the measurements. For this example, all reactions are reversible and so the conditions for EKF failure hold.

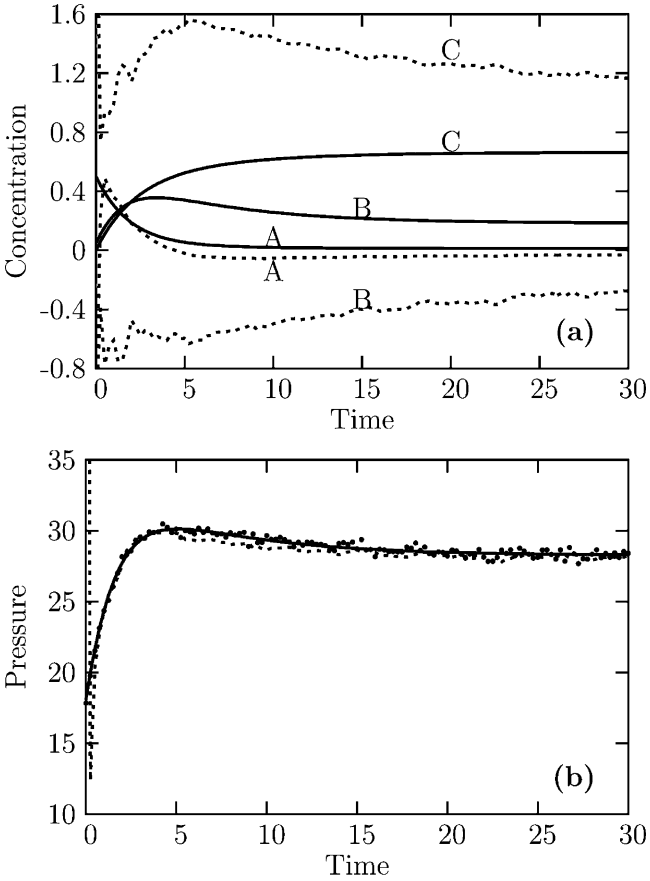

Figure 2. Extended Kalman filter results. (a) Evolution of the actual (solid line) and EKF updated (dashed line) concentrations. (b) Evolution of the actual (solid line), measured (points), and EKF-updated (dashed line) pressure estimates.

$$\begin{gathered}\text{number of linear equations} < \\ \text{number of estimated species} \le \\ \text{number of independent equations} \\ n_m + \eta = 2 < s = 3 \le n + \rho = 3\end{gathered}$$

We now examine the estimation performances of both the EKF and MHE for this system.

Figure 2 demonstrates that the EKF cannot reconstruct the evolution of the state for this system. In fact, the EKF appears to converge to incorrect steady-state estimates of the state. Table 1 presents the results of solving eqs 10 for this system. Note that the concentrations of components A and B are negative, indicating that the EKF has converged to an unphysical state estimate. To prevent negative concentrations, we next implement an ad hoc clipping strategy in which negative filtered values of the state are set to zero (i.e., if $\hat{x}_{k|k} < 0$, set $\hat{x}_{k|k} = 0$). Figure 3 shows a plot of these clipped EKF results. Here, the clipped EKF drives the predicted pressure 3 orders of magnitude larger than the measured pressure before eventually converging to the actual states. Figure 4 presents the results of applying MHE. For these results, we have constrained the state to prevent estimation of negative concentrations. The plots demonstrate that MHE swiftly converges to the correct state estimates.

A little algebraic analysis reveals that multiple steady states satisfy the steady-state measurement for this

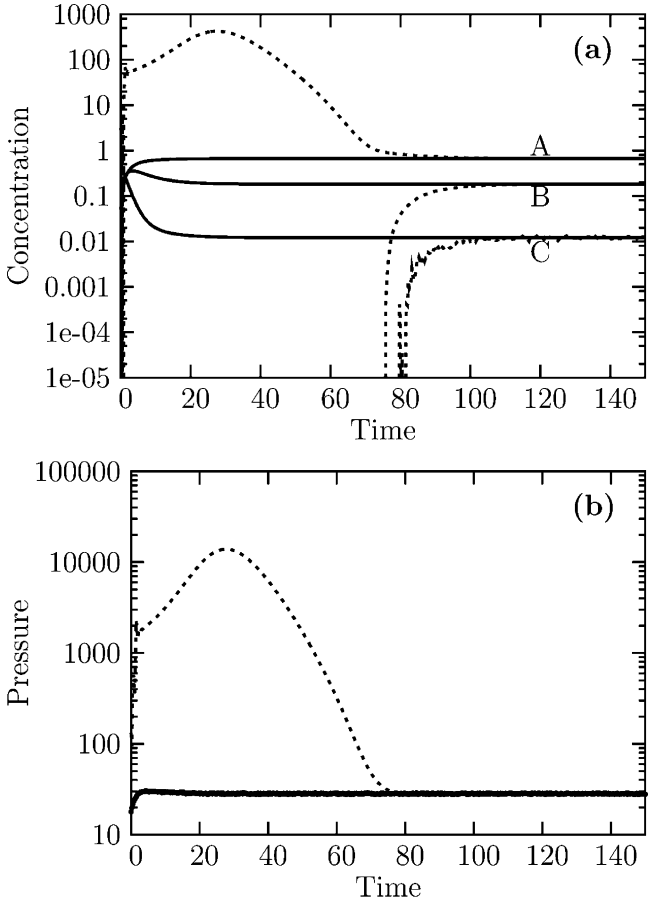

Figure 3. Clipped extended Kalman filter results. (a) Evolution of the actual (solid line) and clipped EKF updated (dashed line) concentrations. (b) Evolution of the actual (solid line), measured (points), and clipped-EKF-updated (dashed line) pressure estimates.

system. At steady state, the model and measurement equations yield one linear equation (assuming no noise in the steady-state measurement y_{ss})

$$c_A + c_B + c_C = \frac{y_{ss}}{RT} \tag{24}$$

and two nonlinear equations (because the rank of the stoichiometric matrix is 2)

$$k_1 c_A = k_{-1} c_B c_C \tag{25}$$

$$k_2 {c_B}^2 = k_{-2} c_C \tag{26}$$

Solving for the steady-state solution using eqs 24−26 gives

$$c_C = \frac{k_2}{k_{-2}} {c_B}^2 = K_2 {c_B}^2 \tag{27}$$

Descartes' rule of signs states that, for polynomials with

$$c_A = \frac{k_{-1}k_2}{k_1 k_{-2}} {c_B}^2 = \frac{K_2}{K_1} {c_B}^3 \tag{28}$$

$$0 = \frac{K_2}{K_1} {c_B}^3 + K_2 {c_B}^2 + c_B - \frac{y_{ss}}{RT} \tag{29}$$

real coefficients, the number of positive, real roots is either the number of sign changes between consecutive

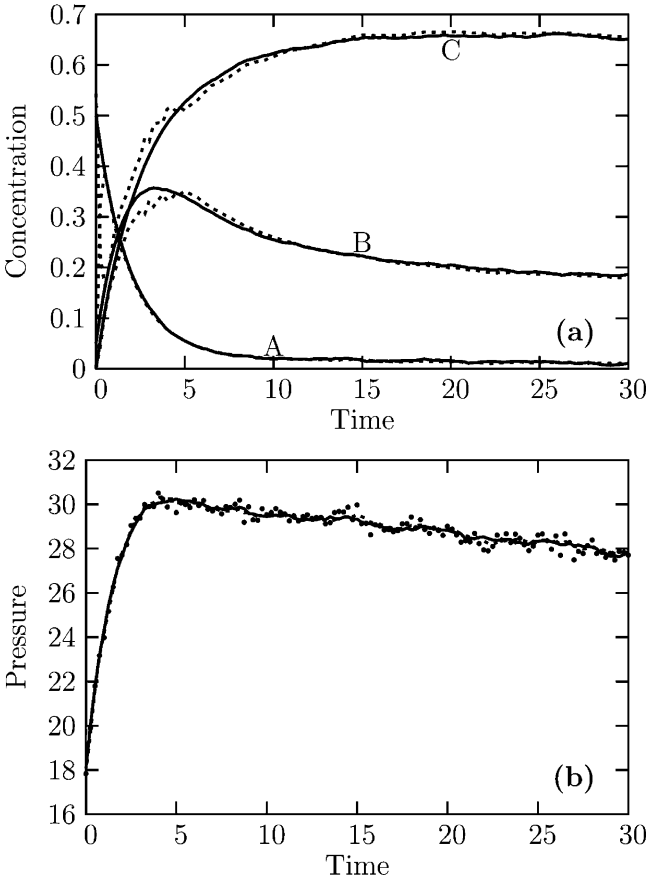

Figure 4. Moving-horizon estimation results, states constrained to $x \geq 0$, smoothing initial covariance update, and horizon length of 2.5 time units ($N = 11$ measurements). (a) Evolution of the actual (solid line) and MHE-updated (dashed line) concentrations. (b) Evolution of the actual (solid line), measured (points), and MHE-updated (dashed line) pressure estimates.

Table 1. EKF Steady-State Behavior, No Measurement or State Noise

component	predicted EKF steady state	actual steady state
A	−0.0122	0.0224
B	−0.1364	0.2006
C	1.1746	0.6411

coefficients or 2 less than this number. Because equilibrium constants and the steady-state measurement are positive, eq 29 has at most one positive root. Thus, there is only one physically realizable steady state. MHE is a natural estimation tool for this system because its incorporation of constraints can prevent the estimator from converging to unphysical steady states.

3.2.2. CSTR. From first principles, the model for a well-mixed, isothermal CSTR reactor is

$$3x = \frac{Q_f}{V_R} c_f - \frac{Q_o}{V_R} x + \nu^T r \tag{30}$$

$$c_f = [0.5 \quad 0.05 \quad 0]^T \tag{31}$$

$$x_0 = [0.5 \quad 0.05 \quad 0]^T \tag{32}$$

$$Q_f = Q_o = 1 \tag{33}$$

$$V_R = 100 \tag{34}$$

We consider state estimation with the following measurement and parameters:

$$y_k = [RT \quad RT \quad RT]x_k \tag{35a}$$

$$\Delta t = t_{k+1} - t_k = 0.25 \tag{35b}$$

$$\Pi_0 = \mathrm{diag}(4^2, 4^2, 4^2) \tag{35c}$$

$$G_k = \mathrm{diag}(1, 1, 1) \tag{35d}$$

$$Q_k = \mathrm{diag}(0.001^2, 0.001^2, 0.001^2) \tag{35e}$$

$$R_k = 0.25^2 \tag{35f}$$

$$\bar{x}_0 = [0 \quad 0 \quad 3.5]^{\mathrm{T}} \tag{35g}$$

The distribution for the initial state is $\mathcal{N}(\bar{x}_0,\Pi_0)$. Again, the initial guess, $\bar{x}_0$, is poor. The actual plant experiences $\mathcal{N}(0,Q_k)$ noise in the state and $\mathcal{N}(0,R_k)$ noise in the measurements. For this example

$$\text{number of linear equations} < \text{number of estimated species}$$

$$n_{\mathrm{m}} + \eta = 2 < s = 3 \tag{36}$$

indicating that multiple measurements might satisfy the steady-state measurement. We now examine the estimation performances of both the EKF and MHE for this system.

Figure 5 demonstrates that, similarly to the batch case, the EKF appears to converge to an incorrect steady-state estimate. This observation is confirmed by determining the EKF steady state assuming no state or measurement noise. Calculating the EKF steady state via eqs 10 and assuming no state or measurement noise yields the results in Table 2. Some steady-state analysis of the system sheds light onto the cause of this phenomenon. Assuming no noise in the steady-state measurement, the system has one linear steady-state measurement y_{ss}

$$c_{\mathrm{A}} + c_{\mathrm{B}} + c_{\mathrm{C}} = \frac{y_{\mathrm{ss}}}{RT} \tag{37}$$

and one linear combination resulting from η_{n}, the null space of the stoichiometric matrix

$$\eta_{\mathrm{n}} = [3 \quad 1 \quad 2] \tag{38}$$

$$3c_{\mathrm{A}} + c_{\mathrm{B}} + 2c_{\mathrm{C}} = 3c_{\mathrm{Af}} + c_{\mathrm{Bf}} + 2c_{\mathrm{Cf}} \tag{39}$$

Therefore the steady-state calculation is a nonlinear problem, and this system satisfies both conditions required for EKF failure.

Figure 6 presents the EKF estimation results for implementation of a clipping strategy. Although clipping eliminates estimation error, this strategy causes a lengthy period of overestimation of the pressure, in some cases by 2 orders of magnitude.

Figure 7 presents the results of applying MHE. For these results, we have constrained the state to prevent estimation of negative concentrations. These figures demonstrate that MHE swiftly converges to the correct state estimates.

3.3. Example 2. Reconsider the batch model given in section 3.2, but with the following updated parameters

$$k = [0.5 \quad 0.4 \quad 0.2 \quad 0.1]^{\mathrm{T}} \tag{40a}$$

$$R_k = 0.1^2 \tag{40b}$$

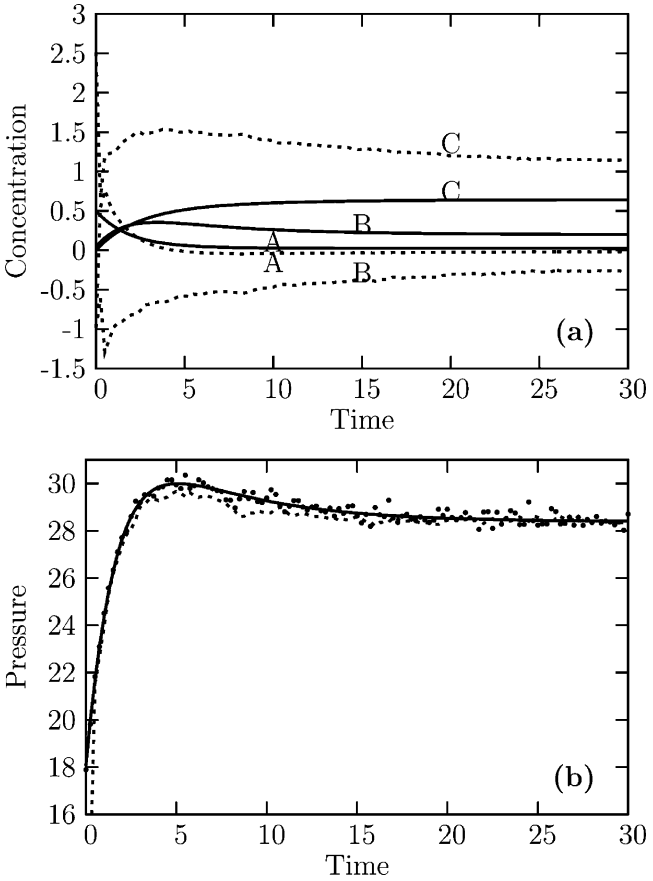

Figure 5. Extended Kalman filter results. (a) Evolution of the actual (solid line) and EKF updated (dashed line) concentrations. (b) Evolution of the actual (solid line), measured (points), and EKF-updated (dashed line) pressure estimates.

Table 2. EKF Steady-State Behavior, No Measurement or State Noise

component	predicted EKF steady state	actual steady state
A	−0.0274	0.0124
B	−0.2393	0.1837
C	1.1450	0.6753

and new measurement

$$y_k = [-1 \quad 1 \quad 1]x_k \tag{41}$$

Although the measurement has no physical meaning, we include this example to illustrate EKF and MHE behavior given multiple physically realizable steady states (in contrast to the previous example, which had only one such state). Clearly, the conditions for EKF failure still hold for this example. Solving for the steady-state solution in terms of c_{B} yields

$$0 = -\frac{K_2}{K_1}{c_{\mathrm{B}}}^3 + K_2{c_{\mathrm{B}}}^2 + c_{\mathrm{B}} - y_{\mathrm{ss}} \tag{42}$$

Again using Descartes' rule of signs and taking into account the specified parameters, eq 42 has two positive roots and one negative root. We now examine the effect of poor initial conditions on the estimation behavior of the EKF and MHE.

Table 3 summarizes the estimation results examined in this section. Given a poor estimate of the initial state $[3\ 0.1\ 3]^{\mathrm{T}}$, the EKF cannot reconstruct the evolution of

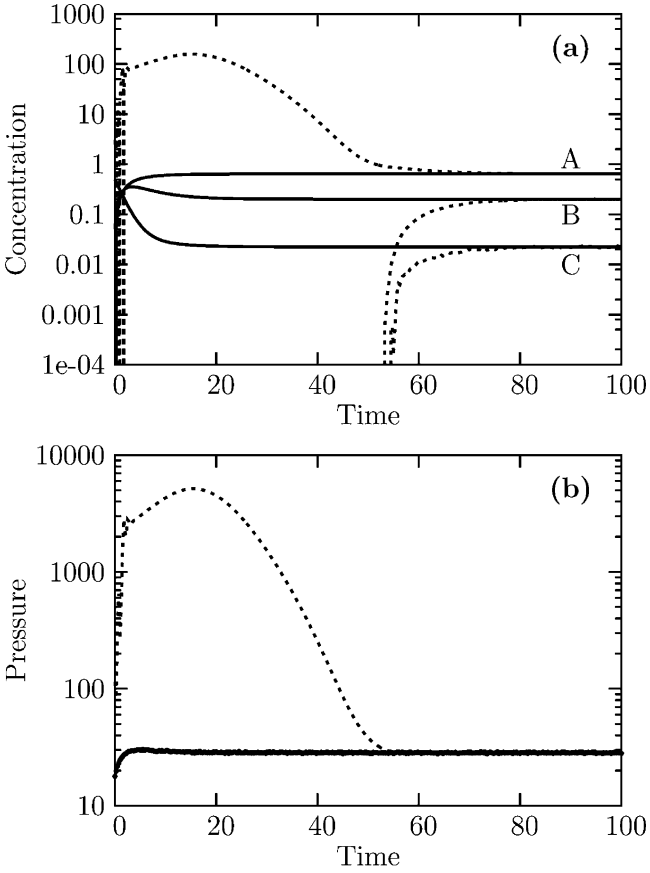

Figure 6. Clipped extended Kalman filter results. (a) Evolution of the actual (solid line) and clipped-EKF-updated (dashed line) concentrations. (b) Evolution of the actual (solid line), measured (points), and clipped-EKF-updated (dashed line) pressure estimates.

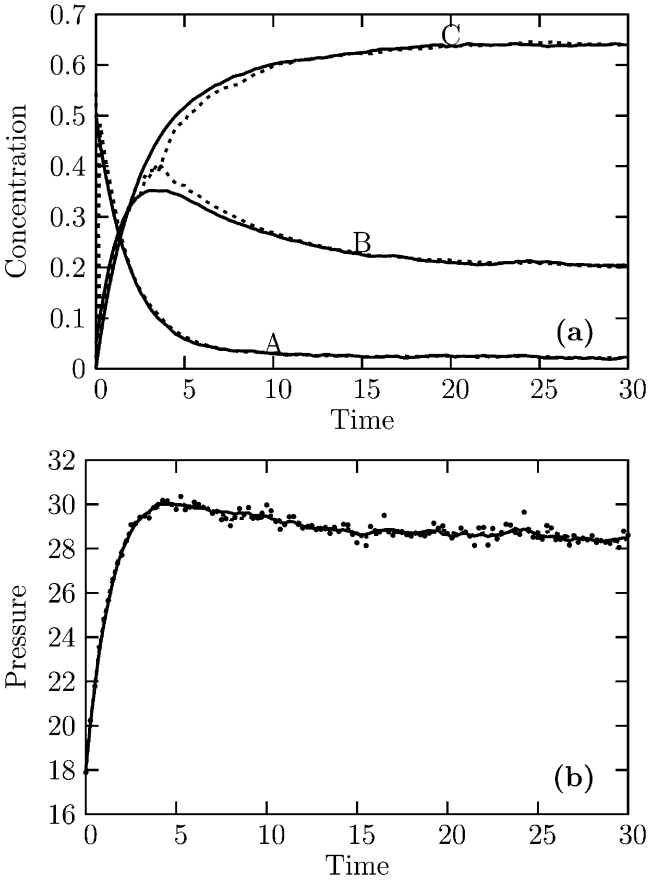

Figure 7. Moving-horizon estimation results, states constrained to $x \geq 0$, smoothing initial covariance update, and horizon length of 2.5 time units ($N = 11$ measurements). (a) Evolution of the actual (solid line) and MHE-updated (dashed line) concentrations. (b) Evolution of the actual (solid line), measured (points), and MHE-updated (dashed line) pressure estimates.

the state, whereas MHE can. Given an even poorer estimate of the initial state $[4\ 0\ 4]^T$, both the EKF and MHE fail to reconstruct the evolution of the state. To improve the quality of the estimates, we constrain the concentrations in the estimators so that

$$0 \leq c_j \leq 4.5, \qquad j = \text{A, B, C} \tag{43}$$

With this extra knowledge, MHE converges to the true state estimates, whereas the clipped EKF estimates are trapped on the constraint. Finally, we relax the concentration constraints to

$$0 \leq c_j \leq 5.5, \qquad j = \text{A, B, C} \tag{44}$$

Not surprisingly, the clipped EKF estimates remain trapped on the constraint. The quality of the MHE estimates is a function of the estimation horizon. If the estimation horizon is too short, the MHE estimates are pinned against the state constraint; increasing the horizon remedies this problem. For short horizons, we suspect that the data in the estimation horizon cannot overcome the biasing of the arrival cost approximation, hence resulting in state estimates pinned against the constraint. Changing arrival cost approximations (e.g., switching from the smoothing scheme to a uniform prior) when constraints are active might constitute one way of addressing this problem without having to increase the estimation horizon.

3.4. Computational Expense. Table 4 summarizes the average computational expense per time step for each of the examples presented in this paper. All computations were performed in GNU Octave (http://www.octave.org/) on a 2.0-GHz processor. MHE computations were performed using the NMPC toolbox (http://www.che.wisc.edu/tenny/nmpc/). Not surprisingly, MHE requires substantially more computational time than the EKF. This increase results because (1) MHE employs optimization whereas the EKF uses a

Table 3. Effects of a Priori Initial Conditions, Constraints, and Horizon Length on State Estimation

estimator	$\bar{x}_0$	constraints	horizon length[a]	estimates converge?[b]
EKF	$[3\ 0.1\ 3]^T$	$x \geq 0$	NA	no
MHE	$[3\ 0.1\ 3]^T$	$x \geq 0$	2.5 time units ($N = 11$)	yes
EKF	$[4\ 0\ 4]^T$	$x \geq 0$	NA	no
MHE	$[4\ 0\ 4]^T$	$x \geq 0$	2.5 time units ($N = 11$)	no
EKF	$[4\ 0\ 4]^T$	$0 \leq x \leq 4.5$	NA	no; CE
MHE	$[4\ 0\ 4]^T$	$0 \leq x \leq 4.5$	2.5 time units ($N = 11$)	yes
EKF	$[4\ 0\ 4]^T$	$0 \leq x \leq 5.5$	NA	no; CE
MHE	$[4\ 0\ 4]^T$	$0 \leq x \leq 5.5$	2.5 time units ($N = 11$)	no; CE
MHE	$[4\ 0\ 4]^T$	$0 \leq x \leq 5.5$	time units ($N = 21$)	no; CE
MHE	$[4\ 0\ 4]^T$	$0 \leq x \leq 5.5$	10 time units ($N = 41$)	yes

[a] N denotes the number of measurements in the estimation horizon. [b] Estimates converge if $x_k - \hat{x}_{k|k} \to 0$. CE indicates that the estimates are trapped on a constraint.

Table 4. Comparison of MHE and EKF Computational Expenses

example	estimator	horizon length[a]	average CPU time per time step (s)
3.2.1	EKF	$N = 1$	0.003
3.2.1	MHE	$N = 11$	0.737
3.2.2	EKF	$N = 1$	0.005
3.2.2	MHE	$N = 11$	0.676
3.3	EKF	$N = 1$	0.006
3.3	MHE	$N = 11$	1.756
3.3	MHE	$N = 21$	4.712
3.3	MHE	$N = 41$	6.899

[a] N denotes the number of measurements in the estimation horizon.

one-step linearization, and (2) MHE calculates sensitivities over a trajectory of states whereas the discrete EKF calculates only a single sensitivity.

4. Conclusions

Virtually all chemical engineering systems contain nonlinear dynamics and/or state constraints. The need to incorporate this information into state estimation is illustrated by the examples presented in this paper. These examples demonstrate that, even with perfect concordance between the model and the physical plant, it is possible for the nominal EKF to fail to converge to the true state when (1) the system model and measurement are such that multiple states satisfy the steady-state measurement and (2) the estimator is given a poor initial guess of the state.

Given the same estimator tuning, model, and measurements as the EKF, MHE provides improved state estimation and greater robustness to poor guesses of the initial state. These benefits arise because MHE incorporates physical state constraints into an optimization, accurately uses the nonlinear model, and optimizes over a trajectory of states and measurements. With local optimization, our results indicate that multivariate normal approximations to the arrival cost combined with judicious use of constraints can prevent multiple optima in the estimator and generate acceptable estimator performance.

One potential pitfall of employing local optimization is the inability to identify multiple modes in the a posteriori distribution. Example 2 of this paper illustrates this pitfall perfectly: attraction of MHE estimates to a mode in the infeasible region leads to state estimates trapped on constraints even though another mode lies within the feasible region. To overcome this difficulty, we are currently investigating the application of Monte Carlo particle filters similar to those proposed by Chen et al.[6] to track multiple optima in the estimator. These filters present one method of identifying the appearance and disappearance of small numbers of local optima in the a posteriori distribution, but they do not provide a reasonable framework for accurately reconstructing the mode of this distribution. Using particle filters to estimate the arrival cost in MHE presents one manner of better approximating the mode of the a posteriori distribution. Additionally, identifying local optima in the arrival cost distribution will yield better initial guesses for the local MHE optimization.

It is reasonable to expect that more complicated models with less restrictive assumptions than the ones proposed here might yield multiple optima corresponding to both physically realizable and unrealizable states. Given that MHE permits incorporation of constraints into its optimization, it is the natural choice for preventing estimation of physically unrealizable states. Because MHE employs a trajectory of measurements as opposed to measurements at only a single time, it is better suited than the EKF for distinguishing among the remaining physically realizable states.

Acknowledgment

The financial support of the National Science Foundation through Grant CTS-0105360 and the industrial members of the Texas−Wisconsin Modeling and Control Consortium is gratefully acknowledged. This material is based on work supported under a National Science Foundation Graduate Research Fellowship. Helpful discussions with Gabriele Pannocchia and Matt Tenny are appreciated as well. We also gratefully acknowledge the organizers of the 2003 Gordon Research Conference for Statistics in Chemistry & Chemical Engineering for providing an environment conducive for discussion of the presented techniques. All simulations were performed using Octave (http://www.octave.org/) and the NMPC toolbox (http://www.che.wisc.edu/tenny/nmpc/). Both Octave and the NMPC toolbox are freely distributed under the terms of the GNU General Public License.

Literature Cited

(1) Soroush, M. State and parameter estimations and their applications in process control. *Comput. Chem. Eng.* **1998**, *23*, 229−245.

(2) Alspach, D. L.; Sorenson, H. W. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Trans. Autom. Control* **1972**, *AC-17*, 439−448.

(3) Gordon, N.; Salmond, D.; Smith, A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F: Radar Signal Process.* **1993**, *140*, 107−113.

(4) Chaves, M.; Sontag, E. State-estimators for chemical reaction networks of Feinberg−Horn−Jackson zero deficiency type. *Euro. J. Control* **2002**, *8*, 343−359.

(5) Bølviken, E.; Acklam, P. J.; Christopherson, N.; Størdal, J.-M. Monte Carlo filters for non-linear state estimation. *Automatica* **2001**, *37*, 177−183.

(6) Chen, W. S.; Ungarala, S.; Bakshi, B.; Goel, P. Bayesian rectification of nonlinear dynamic processes by the weightedbootstrap. Presented at the AIChE Annual Meeting, Reno, NV, Nov 4−9, 2001.

(7) Spall, J. C. Estimation via Markov chain Monte Carlo. *IEEE Control Syst. Mag.* **2003**, *23*, 34−45.

(8) Silverman, B. W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: New York, 1986.

(9) Middlebrooks, S. A. Modelling and Control of Silicon and Germanium Thin Film Chemical Vapor Deposition. Ph.D. Thesis, University of Wisconsin-Madison, Madison, WI, 2001.

(10) Prasad, V.; Schley, M.; Russo, L. P.; Bequette, B. W. Product property and production rate control of styrene polymerization. *J. Process Control* **2002**, *12*, 353−372.

(11) Gudi, R.; Shah, S.; Gray, M. Multirate state and parameter estimation in an antibiotic fermentation with delayed measurements. *Biotechnol. Bioeng.* **1994**, *44*, 1271−1278.

(12) Wilson, D. I.; Agarwal, M.; Rippin, D. Experiences implementing the extended Kalman filter on an industrial batch reactor. *Comput. Chem. Eng.* **1998**, *22*, 1653−1672.

(13) Reif, K.; Günther, S.; Yaz, E.; Unbehauen, R. Stochastic stability of the discrete-time extended Kalman filter. *IEEE Trans. Autom. Control* **1999**, *44*, 714−728.

(14) Reif, K.; Unbehauen, R. The extended Kalman filter as an exponential observer for nonlinear systems. *IEEE Trans. Signal Process.* **1999**, *47*, 2324−2328.

(15) Reif, K.; Günther, S.; Yaz, E.; Unbehauen, R. Stochastic stability of the continuous-time extended Kalman filter. *IEE Proc.: Control Theory Appl.* **2000**, *147*, 45−52.

(16) Stengel, R. F. *Optimal Control and Estimation*; Dover Publications: Mineola, NY, 1994.

(17) Robertson, D. G.; Lee, J. H.; Rawlings, J. B. A moving horizon-based approach for least-squares state estimation. *AIChE J.* **1996**, *42*, 2209−2224.

(18) Rao, C. V.; Rawlings, J. B. Constrained process monitoring: moving-horizon approach. *AIChE J.* **2002**, *48*, 97−109.

(19) Jazwinski, A. H. *Stochastic Processes and Filtering Theory*; Academic Press: New York, 1970.

(20) Tyler, M. L.; Morari, M. *Stability of Constrained Moving Horizon Estimation Schemes*.; Automatic Control Laboratory, Swiss Federal Institute of Technology: Zürich, Switzerland, 1996; Preprint AUT96-18.

(21) Rao, C. V. Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete-Time Systems. Ph.D. Thesis, University of Wisconsin-Madison, Madison, WI, 2000.

(22) Rao, C. V.; Rawlings, J. B.; Mayne, D. Q. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Trans. Autom. Control* **2003**, *48*, 246−258.

(23) Tenny, M. J.; Rawlings, J. B. State Estimation Strategies for Nonlinear Model Predictive Control. Presented at the AIChE Annual Meeting, Reno, NV, Nov 4−9, 2001.

(24) Tenny, M. J.; Rawlings, J. B. Efficient Moving Horizon Estimation and Nonlinear Model Predictive Control. In *Proceedings of the American Control Conference*; IEEE Press: Piscataway, NJ, 2002, pp 4475−4480.

(25) Rao, C. V.; Rawlings, J. B.; Lee, J. H. Constrained linear state estimation—A moving horizon approach. *Automatica* **2001**, *37*, 1619−1628.

(26) Haseltine, E. L.; Rawlings, J. B. A *Critical Evaluation of Extended Kalman Filtering and Moving Horizon Estimation*; Technical Report 2002-03, TWMCC; Department of Chemical Engineering, University of Wisconsin-Madison: Madison, WI, 2002.

Received for review December 15, 2003
Revised manuscript received April 29, 2004
Accepted May 4, 2004

IE034308L