Constrained Particle Filtering Methods for State Estimation of Nonlinear Process

Zhonggai Zhao, Biao Huang, and Fei Liu
Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2V4

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, China, 214122

DOI 10.1002/aic.14390
Published online February 26, 2014 in Wiley Online Library (wileyonlinelibrary.com)

Increasingly in practical applications, nonlinearity, non-Gaussianity, and constraint must be considered to obtain good state estimation. A constrained particle filter (PF) approach for state estimation, which involves three alternative strategies to impose the constraints on the prior particles, posterior particles, and state estimation has been proposed. First, to impose constraints on prior particles, a constrained Gibbs sampling method with a constrained inverse transform sampling is proposed to restrict sampling within the constraint region under cases of both univariate and coupling constraints. Second, to ensure validity of posterior particles, resampling is confined to the valid prior particles and the violated ones are discarded, which results in a similar formulation as the existing acceptance/rejection constrained PF method in literature. Third, if the state estimation violates the constraint, different from the existing methods that either discard all violated particles or accept all of them by projecting them onto the constraint region, the proposed method makes a balance between the prior and the likelihood function by adjusting the weights of violated and valid particles, respectively. Compared with the existing methods, the proposed method provides better physical interpretation and involves no restrictive assumptions about the distributions. Simulation results demonstrate effectiveness of the proposed methods. © 2014 American Institute of Chemical Engineers *AIChE J*, 60: 2072–2082, 2014
Keywords: process control, statistical analysis, constrained particle filter, state estimation, Bayesian method, nonlinear process

Introduction

State estimation plays an important role in many industrial processes. From the Bayesian Statistics view, the objective of state estimation is to infer the posterior probability distribution function (PDF) of state according to the existing measurements. When the models about the process, represented by a state transition equation and a measurement equation, are available, Bayesian methods provide a sequential solution by iteratively predicting states through the state transition equation and updating the estimation using current measurements,[1,2] which means that the PDF of state is sequentially derived through the iterative update of prior distribution using measurements.[3,4]

To achieve mathematical tractability of state estimation, various restrictions are usually made on the system. For instance, if a linear process with Gaussian noise is assumed, Kalman filter (KF) is derived from the Bayesian methods and can obtain optimal solution of state estimation, while for a nonlinear system with Gaussian noise, extended KF (EKF), unscented KF (UKF), and moving horizon estimator (MHE) are often used. Among these methods, KF gives an exact analytical recursive solution for the PDF of state,[5] but the restriction of linearity assumption precludes its applications from nonlinear and non-Gaussian processes. EKF transforms the nonlinear model into a linear one, then the KF procedure is used,[6] but it is only reliable for the mild nonlinear systems. Through unscented transformation to propagate mean and covariance information, UKF is more effective in dealing with nonlinear systems, but its flexibility is restricted by the selection of deterministic sigma points and the assumption of Gaussian distribution.[7] MHE converts the state estimation problem into an optimization formulation in a moving fixed-size window,[8,9] but it typically involves large computation and is also restricted to Gaussian assumption.

Nonlinearity and non-Gaussianity are commonly encountered in real processes. Because of relaxation of the restrictions of linearity and Gaussianity based on the Monte Carlo simulation,[10–12] particle filter (PF) is widely used as an iterative solution to the PDF of states, where particles along with their weights are used to approximate the PDF. In state estimation for both continuous and batch processes, PF has attracted a lot of attention.[13,14]

Conversely, arising from physical restrictions, constraints must be taken into account to ensure the estimation performance. KF and its extensions are not suitable for dealing with

A part of this article was presented at the 15th International Conference on Information Fusion, Singapore, July 2012.

Correspondence concerning this article should be addressed to B. Huang at bhuang@ualberta.ca.

constraints.[15,16] MHE is an attractive approach to handle both nonlinear and linear systems with constraints,[17,18] but it involves excessive computation costs. PF usually adopts two approaches to handle the constraints. One called the acceptance/rejection method imposes constraints on prior particles[19]: Particles satisfying constraints are accepted while those violating constraints are rejected. This method needs no excessive computation and can ensure validity of particles, and, therefore, no violation of estimated states is expected but diversity of particles decreases. Based on the principle of maximizing the posterior, the second approach called the optimization-based method provides a more systematic route to deal with constraints.[20] When a constraint violation occurs, this method computes the valid particle by projecting the violated particle into constraint regions through an optimization formulation. However, some additional assumptions are made on the state or particles to ensure tractability of estimation, such as exponential and double exponential distributions for the prior, and Gaussian distribution is assumed for the posterior PDF to generate posterior particles, which reduces the generality of PF. Moreover, this method is computationally expensive because of the solution of the optimization formulation at every sampling instant and for every violated particle.

This article investigates sampling methods for obtaining constrained particles.[21] Three novel strategies are proposed to impose constraints on prior particles, posterior particles, and state estimation. To ensure no violation of prior particles, the proposed first scheme uses the Gibbs sampling method to obtain the feasible region for each variable, and then, a constrained inverse transform sampling method is proposed to achieve constrained sampling. To obtain valid posterior particles, a constrained resampling scheme is introduced to ensure that the generation of posterior particles is from valid prior particles, which results in a similar formulation as the acceptance/rejection method in literature. To impose the constraint on the state estimation, a novel strategy is proposed to reduce the influence of violated posterior particles and increase that of valid ones by scaling weights of violated and valid posterior particles, respectively. Through employment of the proposed strategies, the constraint problem is solved with better physical interpretation and without assuming Gaussian distribution for posterior.

The remainder of this article is organized as follows. TheBayesian Method for State Estimation section derives the Bayesian method for state estimation. TheParticle Filters section introduces PF methods. TheExisting Constrained PF Method section reviews and analyzes two kinds of the existing constrained PF approaches. In the Proposed Novel Constrained PF Methods section, a constrained Gibbs sampling method, constrained resampling method, and scaling weights method are proposed to deal with the cases of constraint imposed on the prior particles, posterior particles, and ultimate state estimation, respectively. TheSimulations section illustrates the application of the proposed method through numerical simulations. The Conclusions section draws conclusions.

Bayesian Method for State Estimation

Consider a nonlinear state space model as follows

$$\begin{cases} x_k = f_k(x_{k-1}, \omega_{k-1}) \\ y_k = h_k(x_k) + v_k \end{cases} \tag{1}$$

where x_k, y_k, ω_{k-1}, and v_k represent state, output, system noise, and measurement noise at sampling instant k, respectively; f_k, h_k are nonlinear functions at sampling instant k.$p(x_k|Y_{k-1})$ can be derived based on the assumption of the first-order Markov process[4]

$$p(x_k|Y_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Y_{k-1})dx_{k-1} \tag{2}$$

The state prediction $p(x_k|Y_{k-1})$ is determined by the posterior at the previous sampling instant. Conversely, the posterior PDF of states at the current sampling instant is obtained by updating $p(x_k|Y_{k-1})$ with the current measurements, that is, the posterior is the update of the prediction. Considering that the measurements are conditionally independent given states, we have[4,10]

$$p(x_k|Y_k) = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{p(y_k|Y_{k-1})} \tag{3}$$

where $Y_k = \{y_1, y_2, \cdots, y_k\}$ are measurements. Equations 2 and 3 form the recursive solution for the estimation of posterior PDF, which provides a closed-form solution for linear and Gaussian models but is computationally intractable for general nonlinear and non-Gaussian systems. Therefore, PF methods have been proposed.

Particle Filters

To present as well as to understand the proposed three constrained solutions, it is necessary to revisit the generic PF.

Generic PF

To obtain the solution to (2) and (3), Monte Carlo integration is used and the posterior is approximated by[22]

$$\hat{p}(x_k|Y_k) = \frac{1}{N}\sum_{i=1}^{N} \delta(x_k - x_k^i) \tag{4}$$

where x_k^i represents the random sample from the posterior PDF $p(x_k|Y_k)$. However, as the direct sampling from $p(x_k|Y_k)$ is usually intractable, the importance sampling method is widely adopted to indirectly generate particles. A proposal distribution $h(x)$, from which particles are easier to draw, can be used as the importance sampling function if the support of $h(x)$ contains that of $p(x_k|Y_k)$, that is, wherever $p(x_k|Y_k) > 0, h(x) > 0$. Then, the posterior PDF can be approximated by the particles drawn from $h(x)$ along with the corresponding weights.[23,24] The weight of the ith particle is determined by

$$w_k^i = \frac{p(x_k^i|Y_k)}{h(x_k^i)} \tag{5}$$

The PF approach is constructed by combining the sequential importance sampling method with the Bayesian recursive algorithm. Usually, $p(x_k|x_{k-1}^i)$ is chosen as the proposal distribution, and the importance weight is represented by a recursive formulation as[10]

$$w_k^i \propto w_{k-1}^i p(y_k|x_k^i) \tag{6}$$

The posterior PDF $p(x_k|Y_k)$ is then approximated by

$$\hat{p}(x_k|Y_k) = \sum_{i=1}^{N} w_k^i \delta(x_k - x_k^i) \tag{7}$$

However, with increase of measurements, importance weights tend to vary considerably, and eventually all but one weight may degenerate to be negligible. To eliminate the particles with small weights and concentrate on those with large weights, a resampling procedure is usually inserted following the importance sampling steps to reproduce particles with high weights.[25–27] The information of measurements at the current sampling instant is involved in resampling, so the particles before resampling are called prior particles x_k^{i-}, and those after resampling are called posterior particles x_k^i. Posterior particles have the equal weights after the resampling.

The generic PF algorithm can be summarized as follows[17]:

1. Initialization: According to the prior distribution $p(x_0)$, generate initial particles $\{x_0^i\}_1^N$ for sampling instant $k=0$ and set $k=1$.
2. Importance sampling: Generate prior particles $\{x_k^{i-}\}_1^N$ from $p(x_k|x_{k-1}^i)$.
3. Weighting: Calculate the importance weights according to (6) and normalize the weights as $\tilde{w}_k^i = w_k^i / \sum_{i=1}^N w_k^i$.
4. Resampling: Generate posterior particles $\{x_k^i\}_1^N$ through the resampling procedure and reset $\tilde{w}_k^i = 1/N$.
5. Estimation: Estimate the state by $\hat{x}_k = \frac{1}{N}\sum_{i=1}^N x_k^i$, and set $k=k+1$. Go back to step 2.

Although the generic PF has been shown effective in dealing with nonlinear and non-Gaussian state estimation problem, most of the chemical processes are subject to constraints.

The need of constraint estimation: two-state batch reaction example

Consider a two-state batch reaction illustrated by[20,28]

$$2A \xrightarrow{k} B, k=0.16$$

where the reaction rate is $r=kP_a^2$. The partial pressures are the state variables and the measurement variable is defined as the total pressure.

$$x_k = [P_a, P_b]^T, y_k = [1, 1]x_k$$

According to the first principles of the process, the state transition equation is represented as follows

$$\dot{x} = v^T r$$

where $v=[-2, 1]$. The initial states and the sampling interval for discretizing the model are taken as

$$x_0 = [3, 1]^T, \Delta t = 0.1s$$

The system noise follows Gaussian distribution $N([0,0]^T, 10^{-6}I)$, the measurement noise follows $N(0, 10^{-2})$, and the initial guess in the estimation process is $\hat{x}_0 = [0.1, 4.5]^T$ with the covariance matrix diag ([36, 36]). Generic PF is first used for the state estimation, and the results are shown in Figure 1.

Because of the poor initial guess, the estimation of P_b remains negative at the initial stage for a considerable period of time, which is physically impossible. In practice, due to the physical laws, such as material balance, or the physical constraints, such as hard bound on actuators and plant operations, the state should be constrained in a valid region. Therefore, to ensure the validity of estimation, it is necessary to consider constraints in the PF-based estimation method.

Existing Constrained PF Approaches

The state space model with constraints is usually denoted by

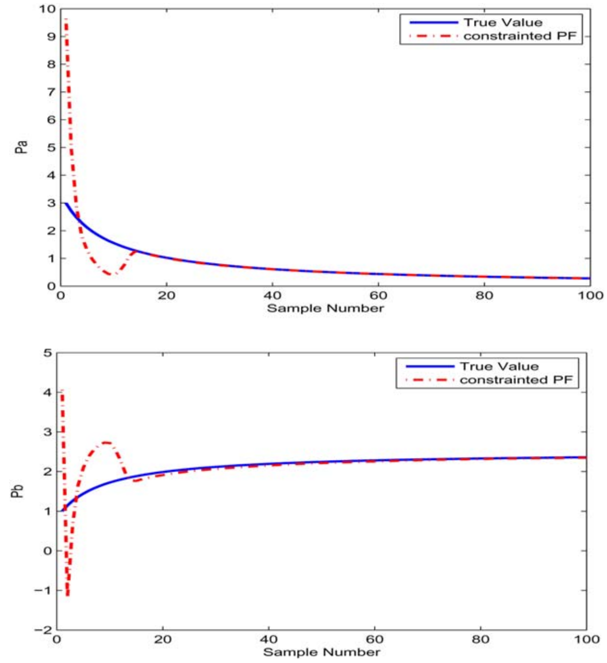

Figure 1. Generic PF estimation, for example, 1.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$\begin{cases} x_k = f_k(x_{k-1}, \omega_{k-1}) \\ y_k = h_k(x_k) + v_k \\ s.t. f(x_k) \in \Omega_k \end{cases} \tag{8}$$

where $f(x_k) \in \Omega_k$ is the constraint. If PF is implemented simply according to the generic PF procedure, the estimation may violate the constraint. There are generally two kinds of methods to handle the constraint problem in PF.

Acceptance/rejection method

This method considers the constraint on the prior particles. The constraint is used as a condition to determine whether the particle should be accepted: those falling inside the constraint region are accepted, while those falling outside the constraint region are rejected. Using the acceptance/rejection method, weights of prior particles are calculated by

$$w_k^i = \begin{cases} 0, & \text{if } (f_k(x_k^{i-}) \notin \Omega_k) \\ \propto p(y_k|x_k^{i-}), & \text{if } (f_k(x_k^{i-}) \in \Omega_k) \end{cases} \tag{9}$$

If some particles lie outside the feasible region, they are rejected and the same number of new particles are produced to satisfy constraints. The new particles along with the valid particles are used to propagate the posterior particles.

The advantage and disadvantage of the acceptance/rejection constrained PF method are obvious. The main advantage is that it can guarantee all particles within the constraint region. As a result, the estimation that is calculated from the particles will also fall inside the constraint. However, it removes all the violated prior particles so the diversity of particles decreases.

Optimization-based method

According to the principle of maximum a posteriori estimation, the state estimation is expressed by

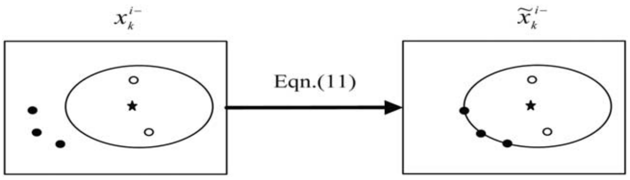

Figure 2. Illustration of the estimation projection: (°: valid particle, •: violated particle, ★: true state).

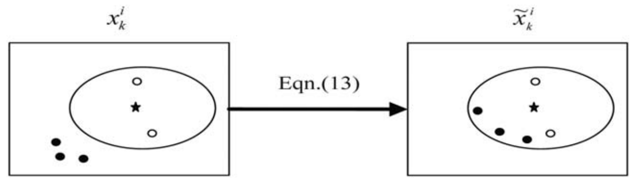

Figure 4. Illustration of the estimation projection: (°: valid particle, •: violated particle, ★: true state).

$$x_k = argmaxp\,(x_k|Y_k) \propto argmaxp\,(x_k|Y_{k-1})p(y_k|x_k) \propto argmaxp_{v_k}(y_k - h_k(x_k^-))p_{x_k^e}(x_k - x_k^-) \quad (10)$$

where p_{v_k} denotes the distribution of measurement noise v_k. As $p(x_k|Y_{k-1})$ is unknown, the assumption $x_k = x_k^- + x_k^e$ is made, where x_k^- is the optimal prediction of x_k through $p(x_k|Y_{k-1})$, and $p_{x_k^e}$ is the distribution of x_k^e approximated by an exponential or double exponential distribution function. Based on these assumptions, constraints can be imposed on prior particles, posterior particles, and the state estimation through optimization formulations as follows[20]

$$\min_{\tilde{x}_k^{i-}} -\log\left(p_{x_k^e}(\tilde{x}_k^{i-} - x_k^{i-})\right) \quad (11)$$

$$\min_{\tilde{x}_k^{i-}} -\log\left(p_{x_k^e}(\tilde{x}_k^{i-} - x_k^{i-})\right) - \log\left(p_{v_k}(y_k - h_k(\tilde{x}_k^{i-}))\right) \quad (12)$$

$$\min_{\tilde{x}_k^{i}} -\log\left(p_{x_k^e}(\tilde{x}_k^{i} - x_k^{i})\right) - \log\left(p_{v_k}(y_k - h_k(\tilde{x}_k^{i}))\right) \quad (13)$$

$$\min_{\tilde{x}_k} -\log\left(p_{x_k^e}(\tilde{x}_k - \hat{x}_k)\right) - \log\left(p_{v_k}(y_k - h_k(\tilde{x}_k))\right) \quad (14)$$

When a particle or state estimation violates the constraint, the optimization procedure finds a projected value within the constraint region in place of the violated particle or estimation rather than simply discarding them, which can overcome the demerits of the acceptance/rejection method.

The first-optimization scheme denoted by (11) is direct and simple, but it only considers minimizing deviation between the projected and the violated particles, so the projection of the violated particles will always fall on the boundary of constraint regions as shown in Figure 2, where the rectangle represents the state space and the ellipse the constraint region. For instance, if the constraint region is $x_k \geq 0$, the particles that are less than 0 are all set to 0. As a result, the estimation tends to be close to the boundary if there are significant particles violating the constraints. Furthermore, this approach can significantly change the distribution of particles and eventually affect the state estimation.

Similarly, imposing constraints on the prior particles, the second-optimization scheme makes a trade-off between the particle deviation and output deviation by (12). This method considers the influence of outputs when drawing back the violated prior particles into the constraint region, so the projections may not fall on the boundary of constraint regions according to Figure 3.

In the third-optimization method, the violation of prior particles is permitted and resampling is implemented among all the prior particles. It aims at the validity of posterior particles by using (13). Both output and particle deviations are taken into account, when projecting the violated posterior particles into the feasible region. However, every violated particle needs to be handled to ensure the validity by optimization algorithms as shown in Figure 4, which will result in expensive computation.

The fourth-optimization method aims at imposing constraints directly on the estimation rather than on particles. The violated estimation will be projected into the constraint region by (14). However, the posterior particles are regenerated according to the estimated state based on the assumption that the posterior particles are normally distributed around the estimated state as shown in Figure 5, which may violate the real distribution.

Proposed Novel Constrained PF Methods

Following the idea in the optimization-based method that the constraints are imposed on the prior particles, the posterior particles, and the state estimation, this article considers the constraints in the PF method in the same way and propose a constrained PF approach to overcome shortcomings of the optimization-based method, consisting of three schemes imposing constraints on prior particles, posterior particles, and state estimation.

Novel constrained sampling for prior particles

From the generic PF procedure, $p(x_k|x_{k-1}^i)$ is usually chosen as the importance sampling function to produce prior particle i. So the constraint on prior particles is converted into how to draw particles from $p(x_k|x_{k-1}^i)$ within the constraint region $f(x_k) \in \Omega_k$. In the case of univariate constraint and coupling constraint, this section proposes a novel constrained sampling method by using inverse sampling and Gibbs sampling strategies.

Univariate Constraint. The particle x_k^{i-} is produced by substituting the previous posterior particle x_{k-1}^i and the noise particle w_{k-1}^i into the state transition equation as

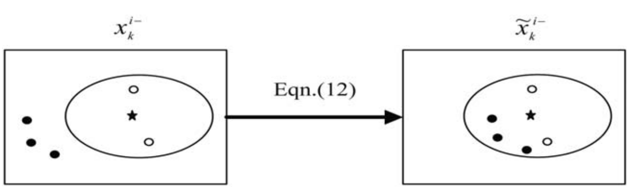

Figure 3. Illustration of the estimation projection: (°: valid particle, •: violated particle, ★: true state).

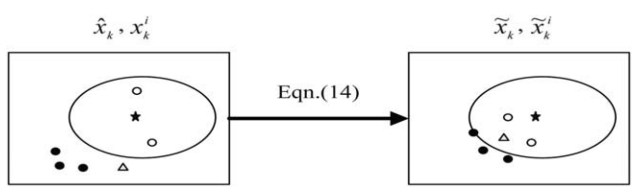

Figure 5. Illustration of the estimation projection: (°: valid particle, •: violated particle, △: estimation, ★: true state).

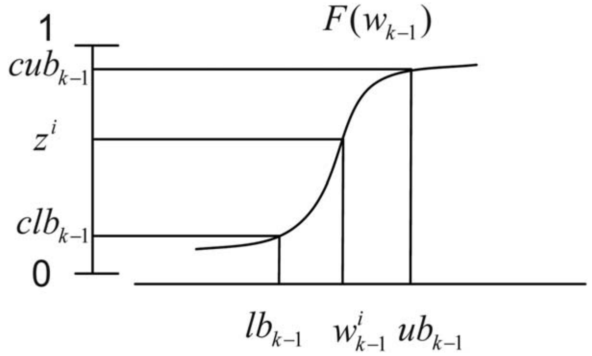

Figure 6. Illustration of constrained inverse transform sampling.

$$x_k^{i-}=f_k(x_{k-1}^i, w_{k-1}^i) \in \Omega_k \tag{15}$$

where x_{k-1}^i is a deterministic term at sampling instant k, and w_{k-1}^i is the uncertain term which determines the validity of x_k^{i-}. To ensure the validity of prior particles, sampling of w_{k-1}^i must be limited to the valid range of system noise calculated by (15), and the valid region is denoted by

$$lb_{k-1} \leq \omega_{k-1}^i \leq ub_{k-1} \tag{16}$$

Letting $F(\omega_{k-1})$ be the cumulative density function (CDF) of $p(\omega_{k-1})$, the range of cumulative probability for $\omega_{k-1}, \bar{\omega}_{k-1}$, is derived as

$$clb_{k-1} \leq \bar{\omega}_{k-1} \leq cub_{k-1} \tag{17}$$

where clb_{k-1} and cub_{k-1} are the cumulative probabilities corresponding to the range value lb_{k-1} and ub_{k-1} on the PDF, respectively.

Given the range of the system noise and its cumulative probability, Figure 6 illustrates the proposed constrained inverse transform sampling. Uniformly extract a sample z^i from the range of $[clb_{k-1}, cub_{k-1}]$ and then reversely calculate the value on the density function of system noise corresponding to $z^i, \hat{\omega}_{k-1}^i=F^{-1}(z^i)$, which is regarded as a sample from $p(\omega_{k-1})$.[29] Then, the prior particle is expressed by $x_k^{i-}=f_k(x_{k-1}^i, \hat{\omega}_{k-1}^i)$.

From the above procedure, the generation of each constrained prior particle requires four steps as follows:

1. Calculate the feasible range of system noise according to the state transition equation and the state constraint, $lb_{k-1} \leq \omega_{k-1} \leq ub_{k-1}$.
2. Convert the feasible range of density function into that of cumulative probability, $clb_{k-1} \leq \bar{\omega}_{k-1} \leq cub_{k-1}$.
3. Draw a sample z^i from uniform $[clb_{k-1}, cub_{k-1}]$ distribution.
4. Reversely transform the extracted probability into a random sample value, $\hat{\omega}_{k-1}^i=F^{-1}(z^i)$.

To reduce the computation load, "3σ" rule is used to determine whether it is necessary to perform all steps. In Figure 7, ω_m is the mean value of ω_{k-1}. If the range $[lb_{k-1}, ub_{k-1}]$ encloses the range $[\omega_m-3\sigma, \omega_m+3\sigma]$ within it, almost all particles will fall inside the constraint region, which will result in valid state estimation. Therefore, in this case, the last three steps can be skipped and particles are directly drawn from the distribution of the random term with no violation of estimation.

This scheme transfers the constraint on particle to that on the system noise, and avoids constraint violation of prior particles before it occurs. By substituting the previous posterior particle into the state transition equation, the constraints of density function and then CDF of system noise are calculated. The prior particles within feasible region are obtained by uniformly sampling within the valid range of CDF. The constrained inverse transform sampling needs no optimization calculation and no assumption about the distribution and provides an efficient way to generate valid prior particles.

The proposed constrained inverse transform sampling method is suitable for the PDF with easily obtained inverse CDF. Moreover, this method can deal with the uncoupling multivariate constraint which can convert into univariate one, but the coupling constraint is beyond the capacity of this method.

Coupling Constraint. Given an N_x-dimensional state vector with coupling constraint, it is difficult to implement the constrained inverse transform sampling as discussed in the previous section. Considering that the validity of particles can be ensured by making each variable within the constraint region sequentially, the constrained sampling method based on Gibbs sampling method is proposed in this section: Gibbs sampling method is first used to obtain the constraint region of each variable, and then the constrained inverse transform sampling method discussed in the previous section is used to generate valid particles.

At sampling instant k, the goal of sampling is to draw particles from $p(x_k|x_{k-1}^i)$ within the constraint region $f(x_k) \in \Omega_k$. Assuming that the starting values of particles are $x_{1,k}^{(0)}, x_{2,k}^{(0)}, \cdots, x_{N_x,k}^{(0)}$ and $m=1$, random values are obtained by the following steps.

1. Given $x_{2,k}^{(m-1)}, x_{3,k}^{(m-1)}, \cdots, x_{N_x,k}^{(m-1)}$, the constraint region of $x_{1,k}, \Omega_{1,k}^{(m)}$, is calculated according to $f_k(x_k) \in \Omega_k$.
2. Sample $x_{1,k}^{(m)}$ from $p(x_{1,k}|x_{k-1}^i, x_{2,k}^{(m-1)}, x_{3,k}^{(m-1)}, \cdots, x_{N_x,k}^{(m-1)})$ within the constraint region $x_{1,k} \in \Omega_{1,k}^{(m)}$.
3. The constraint region of $x_{2,k}^{(m)}, \Omega_{2,k}^{(m)}$, is calculated given $x_{1,k}^{(m)}, x_{3,k}^{(m-1)}, x_{4,k}^{(m-1)}, \cdots, x_{N_x,k}^{(m-1)}$ and $f_k(x_k) \in \Omega_k$.
4. Sample $x_{2,k}^{(m)}$ from $p(x_{2,k}|x_{k-1}^i, x_{1,k}^{(m)}, x_{3,k}^{(m-1)}, x_{4,k}^{(m-1)}, \cdots, x_{N_x,k}^{(m-1)})$ within the constraint region $x_{2,k} \in \Omega_{2,k}^{(m)}$.
5. Similarly, $x_{3,k}^{(m)}, x_{4,k}^{(m)}, \cdots, x_{N_x,k}^{(m)}$ are generated in sequence.
6. Set $m=m+1$, iteratively implement steps from one to five.

After $m=h$ iterations, the Gibbs sampler will converge to $p(x_k|x_{k-1}^i)$ in the constraint region,[30] and $x_{1,k}^{(m)}, x_{2,k}^{(m)}, \cdots, x_{N_x,k}^{(m)}$ are regarded as the realization of x_k^{i-}.

From the constrained Gibbs sampling, sampling x_k^{i-} from $p(x_k|x_{k-1}^i)$ in $f(x_k) \in \Omega_k$ is converted to sampling $x_{j,k}^{(m)}$ from $p(x_{j,k}|x_{k-1}^i, x_{1,k}^{(m)}, x_{2,k}^{(m)}, \cdots, x_{j-1,k}^{(m)}, x_{j+1,k}^{(m-1)}, x_{j+2,k}^{(m-1)}, \cdots, x_{N_x,k}^{(m-1)})$ in $x_{j,k} \in \Omega_{j,k}^{(m)}$, which belongs to a univariate case. According to the previous section, the constraint of the jth element on its

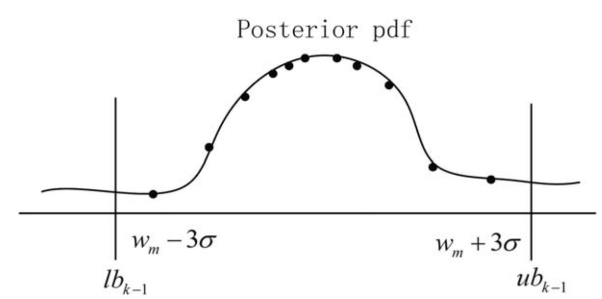

Figure 7. Illustration of 3σ of system noise.

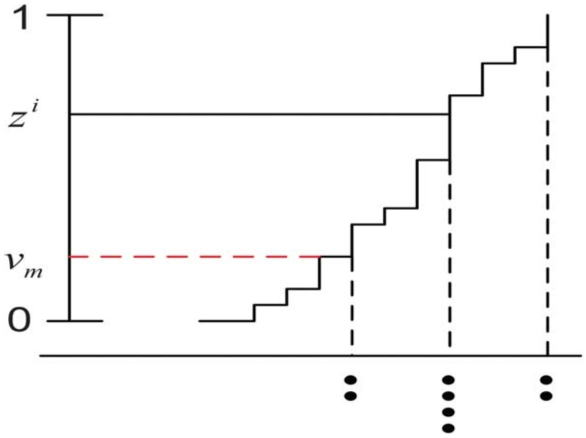

Figure 8. Illustration of constrained resampling.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

CDF is calculated. Uniformly draw a sample from the constraint region on its CDF, and then inversely transform the sample into the density function to obtain the random particle. If it is difficult to get the CDF and inverse transformation, the Griddy-Gibbs sampling method can also be considered.[31]

Novel constrained sampling for posterior particles

Let the prior particles x_k^{i-} be generated without constraint, and without loss of generality assume that $\{x_k^{1-}, x_k^{2-}, \cdots, x_k^{m-}\}$ are violated particles. The sum of normalized weights of the violated prior particles is $\tilde{w}_k^1+\tilde{w}_k^2+\cdots+\tilde{w}_k^m=\pi_v$. To produce the valid posterior particles from the prior particles, the constrained resampling is proposed and illustrated in Figure 8, where the uniform sampling is restricted to the range $[\pi_v, 1]$, that is, the resampling is implemented among the valid prior particles and the violated prior particles are discarded during the resampling stage. Therefore, the essence of the proposed constrained resampling method is the same as that of the acceptance/rejection constrained PF in literature. However, the constrained resampling method proposed here accepts or rejects the prior particles to ensure the validity of the posterior particles, while the objective of acceptance/rejection constrained PF in literature is to ensure the prior particles to be generated from the constraint region. The proposed scheme avoids constraint violations before they occur, and the acceptance/rejection constrained PF in literature is a passive method to recalculate the estimation after it violates the constraint.

Novel scheme for constraint on state estimation

The third way to impose constraint is to put constraint on the ultimate state estimation. Considering that the prior particles and posterior particles are generated without constraints, and assuming that $x_k^1, x_k^2, \cdots, x_k^m$ are the violated posterior particles, the posterior PDF is approximated by

$$\hat{p}(x_k|Y_k)=\frac{1}{N}\sum_{i=1}^{N}\delta(x_k-x_k^i) =\frac{1}{N}\sum_{i=1}^{m}\delta(x_k-x_k^i)+\frac{1}{N}\sum_{i=m+1}^{N}\delta(x_k-x_k^i) \tag{18}$$

The state is estimated by the mean of posterior particles as

$$\hat{x}_k=\frac{1}{N}\sum_{i=1}^{N}x_k^i \tag{19}$$

If the violated posterior particles play a more important role in the calculation of state estimation than the valid ones, the state estimation $\hat{x}_k$ will likely violate the constraint. A strategy to adjust the influence of particles on the state estimation is proposed by scaling the weights of violated as well as valid posterior particles in this section. To avoid confusion, the scale of weights is called coefficient in the sequel.

To ensure the state estimation within the constraint, the influence of violated posterior particle should be decreased and that of valid ones should be increased by scaling their weights. Assuming that the coefficients are α and β, the weights of posterior particles should then be expressed by

$$w_k^i=\begin{cases}\alpha*\dfrac{1}{N}, & \text{if } (f_k(x_k^i)\notin\Omega_k)\\[2ex] \beta*\dfrac{1}{N}, & \text{if } (f_k(x_k^i)\in\Omega_k)\end{cases} \tag{20}$$

where the unity sum of weights should be satisfied

$$\alpha*\frac{1}{N}*m+\beta*\frac{1}{N}*(N-m)=1 \tag{21}$$

so

$$\beta=\frac{N-m\alpha}{N-m} \tag{22}$$

Therefore, the corrected state estimation $\tilde{x}_k$ is written as

$$\tilde{x}_k=\frac{\alpha}{N}\sum_{i=1}^{m}x_k^i+\frac{1-\frac{\alpha m}{N}}{N-m}\sum_{i=m+1}^{N}x_k^i \tag{23}$$

Similar to the fourth strategy of optimization-based methods as discussed in the Existing Constrained PF Method Section, considering the trade-off between the prior and the likelihood function, α can be calculated by the following optimization

$$\begin{aligned}\hat{\alpha}=\min{}_{\hat{\alpha}}\ &-\log\left(p_{v_k}(y_k-h_k(\tilde{x}_k))\right)-\log\left(p_{x_k^e}(\tilde{x}_k-\hat{x}_k)\right)\\ s.t.\ \tilde{x}_k&=\frac{\alpha}{N}\sum_{i=1}^{m}x_k^i+\frac{1-\dfrac{\alpha m}{N}}{N-m}\sum_{i=m+1}^{N}x_k^i\\ f_k(\tilde{x}_k)&\in\Omega_k\\ 0\le\ &\alpha\le 1\end{aligned} \tag{24}$$

Different from (14), only the coefficient of weights is calculated by (24), and the posterior particles remain unchanged and are used for reproduction after the correction of state estimation, so the diversity of particles does not decrease. Conversely, the weights of all valid particles increase in the same scale, so the distribution of neither valid posterior particles nor the violated posterior particles does change.

Using the scaling idea proposed here, we can look back to the acceptance/rejection method and the proposed constrained resampling method in the previous section. From (9) and Figure 7, the two methods accept the valid particles and reject the violated ones, which means the weights of particles are represented by

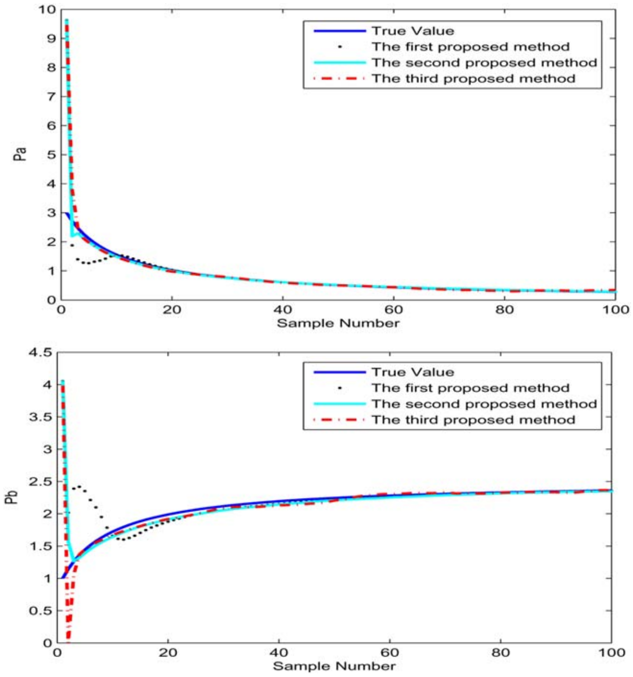

Figure 9. Estimation by the proposed methods, for example, 1.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$w_k^i = \begin{cases} \propto 0 * p(y_k|x_k^{i-}), & \text{if } (f_k(x_k^{i-}) \notin \Omega_k) \\ \propto 1 * p(y_k|x_k^{i-}), & \text{if } (f_k(x_k^{i-}) \in \Omega_k) \end{cases} \tag{25}$$

In other words, the weights of valid particles and violated particles are scaled by "1" and "0," respectively, so both the original acceptance/rejection method and the proposed constrained resampling method can be regarded as a special form of the proposed scaling method. Originating from the acceptance/rejection method in literature and the optimization-based method, the proposed method scales the weights of posterior particles rather than simply discards all violated particles and keeps the posterior particles unchanged rather than reproducing from the state estimation with an assumption of Gaussian distribution, so this method overcomes many disadvantages of the two methods alone.

Simulations

In this section, two cases are used for the evaluation of the proposed method. In the first case, the constrained PF is compared with the generic PF in the state estimation of a two-state batch process. The state estimation for penicillin fermentation process is illustrated in the second case study, and the optimization-constrained PF methods are used for the comparison with the proposed method. All the algorithms are run on a 2.9 GHz CPU with 8 GB RAM PC using MATLAB 2012a.

Two-state batch reaction

The previous motivation section has provided the model description and the lower constraint limit of states. The upper constraint limit of states is set to $[10, 10]^T$, and the initial guess for states is assumed to be $[9.64, 4.06]^T$. Figure 9 illustrates the estimation results by the proposed constrained PF method.

If the initial states are guessed as $[6, 0.8]^T$, the estimation results are shown in Figure 10.

Comparing Figure 1 with Figures 9 and 10, the estimations by the generic PF can roughly follow the real states, but the performance of this method is the worst among all the PF methods. With the poor initial guess for the states, the generic PF may result in the violation of state estimations against their constraints, but through the proposed constrained PF approach, the validity of estimation can always be guaranteed and better estimations are obtained. Similar to the generic PF, the estimations by the proposed methods approach to their true values when more measurements available.

Penicillin fermentation process

Model Description. Penicillin fermentation process is a typical nonlinear and non-Gaussian process with complex mechanism. In the fermenter, the environmental factors for the growth of biomass are constituted by pH value, temperature, dissolved oxygen. To guarantee the ideal growth environment, these variables should be controlled by the addition of acid and base, the circulations of cold water and hot water, and adjustments of air inlet and agitation power, respectively. To provide a public plant for researchers to test novel methods, a simulation model has been developed for penicillin fermentation process by Birol and coworkers as[32,33]

$$\frac{dC_X}{dt} = \mu C_X - \frac{C_X}{V}\frac{dV}{dt} \tag{26}$$

$$\frac{dC_P}{dt} = \mu_{pp} C_X - K C_P - \frac{C_P}{V}\frac{dV}{dt} \tag{27}$$

$$\frac{dC_S}{dt} = -\frac{\mu}{Y_{x/s}} C_X - \frac{\mu_{pp}}{Y_{p/s}} C_X - m_x C_X + F - \frac{C_S}{V}\frac{dV}{dt} \tag{28}$$

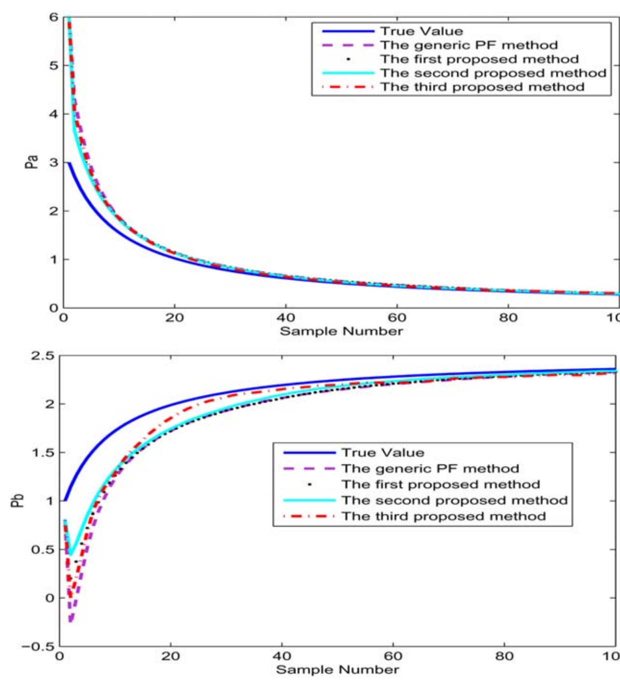

Figure 10. Estimation by the proposed methods, for example, 1.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. True Initial Parameters

State	Description	Initial Value
C_X	Biomass concentration (g/L)	0.1
C_P	Penicillin concentration (g/L)	0
C_S	Substrate concentration (g/L)	15
D_o	Dissolved oxygen (g/L)	1.16
V	Culture volume (L)	100
CO_2	CO_2 evolution (mmol/L)	0.5

$$\frac{dD_o}{dt}=-\frac{\mu}{Y_{x/o}}C_X-\frac{\mu_{pp}}{Y_{p/o}}C_X-m_oC_X+K_{la}(D_{o\max}-D_o)-\frac{C_S}{V}\frac{dV}{dt} \tag{29}$$

$$\frac{dV}{dt}=F-F_{\rm loss} \tag{30}$$

$$\frac{dCO_2}{dt}=\alpha_1\frac{dC_X}{dt}+\alpha_2C_X+\alpha_3 \tag{31}$$

where C_S, C_X, C_P, D_O, V, and CO_2 denote substrate concentration, biomass concentration, penicillin concentration, dissolved oxygen, volume, and CO_2 evolution, respectively. μ and μ_{pp} are the substrate-to-biomass conversion rate and the specific penicillin production rate, respectively, represented by

$$\mu=\mu_{\max}\frac{C_S}{K_cC_X+C_S}\frac{D_o}{K_{ox}C_X+D_o} \tag{32}$$

$$\mu_{pp}=\mu_p\frac{C_S}{K_p+C_S+C_S^2/K_I}\frac{D_o^p}{K_{op}C_X+D_o^p} \tag{33}$$

where $\mu_{\max}$ is the maximum specific growth rate, and $\mu_p, K_p, K_I, p, K_{op}, Y_{x/s}, Y_{p/s} m_x, Y_{x/o}, Y_{p/o}, m_o, D_{o\max}$, and K_{la} are constant parameters. F is the input of fermenter that will change its volume, and $F_{\rm loss}$ is the evaporative loss of fermenter denoted by

$$F_{\rm loss}=V\lambda(e^{5((T-T_0)/T_v-T_0)}-1) \tag{34}$$

where T_0 and T_v are the freezing and boiling temperature, and λ the evaporative rate.

According to the model description, C_X, C_S, and C_P are the states to be estimated, and the measurements are D_o, V, and CO_2, which are available accurately on-line. Discrete-time transformation is implemented to obtain the discretized model.

Simulation Results. The true initial states and parameters are set as shown in Table 1 and the values of constant parameters in the model can be found in the literature.[32,33] The biomass grows in the fermenter starting from the initial biomass concentration, substrate concentration and penicillin concentration 0.1, 15, and 0 g/L, respectively. The sample interval and simulation duration are assumed to be 0.01 and 80 h, respectively. The non-negative states, the biomass concentration, the penicillin concentration, and the substrate concentration, are limited to be less than 10, 1, and 16 g/L, respectively.

The system noise ω_k is assumed to be the absolute value of a random vector that follows a Gaussian distribution $N([0,0,0]^T, \text{diag}([1e{-}3, 1e{-}4, 1e{-}3]))$, and the measurement noise v_k follows the distribution $N([0,0,0]^T, \text{diag}([1e{-}2, 1e{-}2, 1e{-}2]))$. A poor initial state guess is taken as $[0.5, 0, 14.5]^T$ with noise of Gaussian distribution $N([0,0,0]^T, \text{diag}([1e{-}1, 1e{-}2, 1e{-}1]))$, and the accurate initial values for the measurement variables $[1.18, 99.9, 0.49]^T$ because of accurate detection of these variables in practice. A total of 100 particles are used for the Monte Carlo simulations. The estimation result by the generic PF is shown in Figure 11, and the elapsed time for the state estimation of a whole batch is 100.25 s.

As no constraint is imposed on the estimation and the initial guess for state is poor, in the initial fermentation stage, the estimations for product concentration consistently violate the constraint region. Now, the constrained PF is used, and the elapsed time for the state estimation is shown in Table 2.

If the constraint is imposed on the prior, the proposed method should compute the constraint region of system noise for every particle at each sampling instant, and then uniformly extract a sample followed by inverse transformation, so the calculation load is larger than the first optimization method shown in (12) to obtain the valid prior particles. For the second proposed method, no optimization calculation is involved to guarantee the validity of posterior particles, so the second proposed method takes less time for the state estimation than the second optimization method shown in (13). As for the third proposed method, only the coefficient is optimized and no posterior particles need to be regenerated if the estimation violates the constraint, so the third proposed

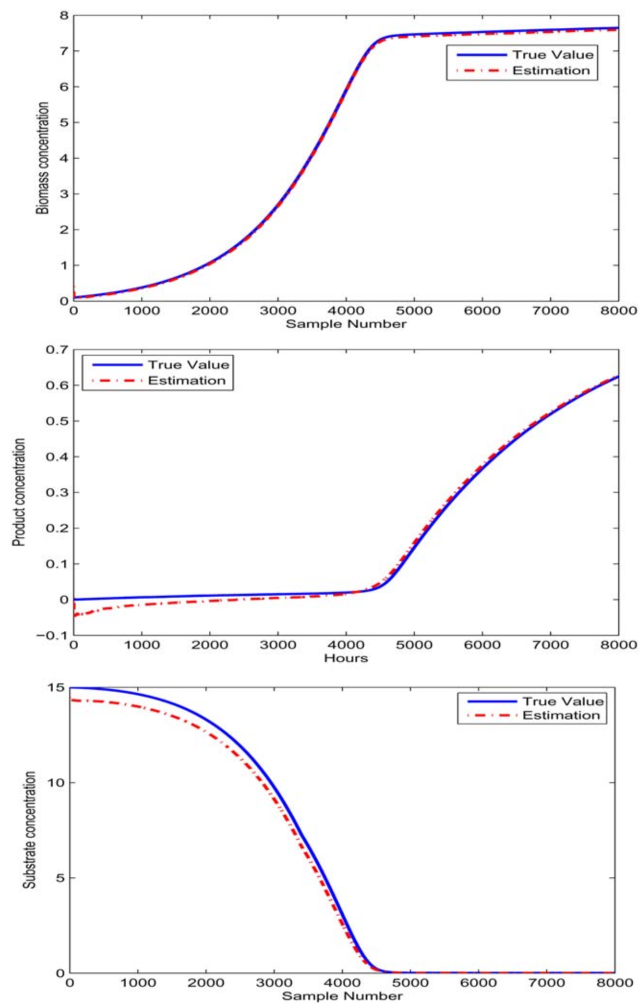

Figure 11. Estimation through the generic PF, for example, 2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 2. Elapsed Time for Estimation Through Constrained PF Algorithms (unit: s)

Method	Constraint on Prior	Constraint on Posterior	Constraint on Estimation
Optimization formulation	336.353276	105.546958	121.444452
Proposed method	426.179965	102.023205	108.897357

method performs better in the calculation load if the constraint is imposed on the estimation.

Considering that the constraint is imposed on the prior particles, posterior particles, and state estimation, the proposed methods and optimization-based methods are compared and the estimation results are shown in Figures 12, 13, and 14.

In Figure 12, the estimation results by the first proposed method are similar to those by the first optimization method. However, in Figure 13, the estimation of product concentration by the second proposed method is slightly worse than that by the second optimization method, which may be caused by the fact that the second proposed method accepts the prior particles only according to the state constraint and ignores the influence of outputs, but the second optimization method shown in (13) obtains the projected particles by considering both the particle deviation and output deviation. The drawback of the second proposed method imposing constraint on posterior particles is the same as that of the acceptance/rejection method in literature. In Figure 14, as the third optimization method shown in (14) should regenerate posterior particles based on the assumption that the posterior particles are normally distributed around the estimation, which may violate the true distribution of posterior particles, the method even performs worse than the generic PF method except for the nonviolation case. Conversely, the third proposed method only need adjusting the weights of each particle to ensure the estimation to be fallen into the constraint region, so it can obtain much better estimation results.

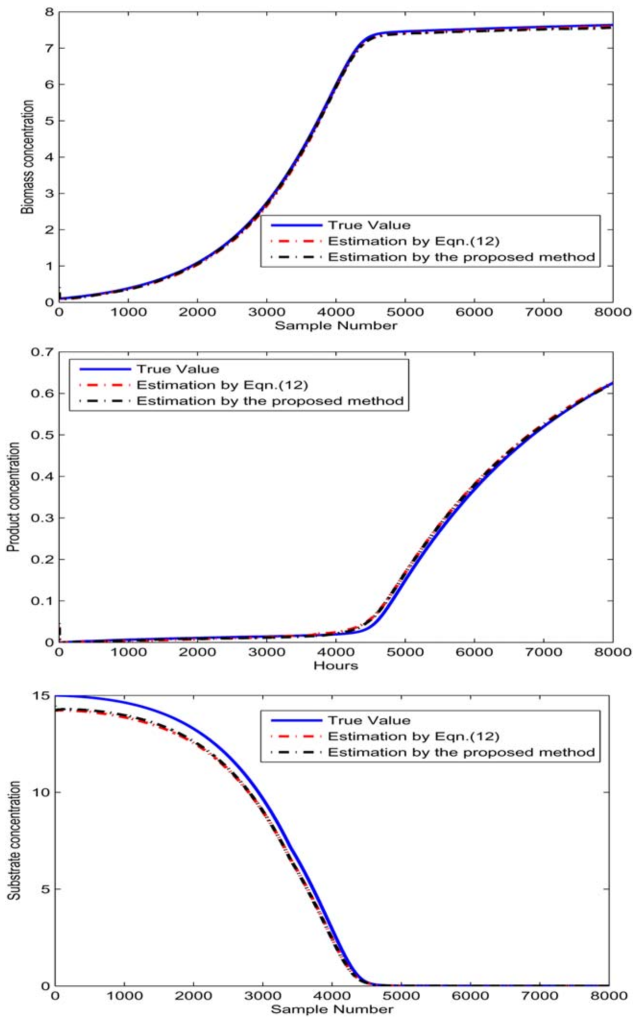

Figure 12. Estimation through imposing constraint on prior particles, for example, 2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

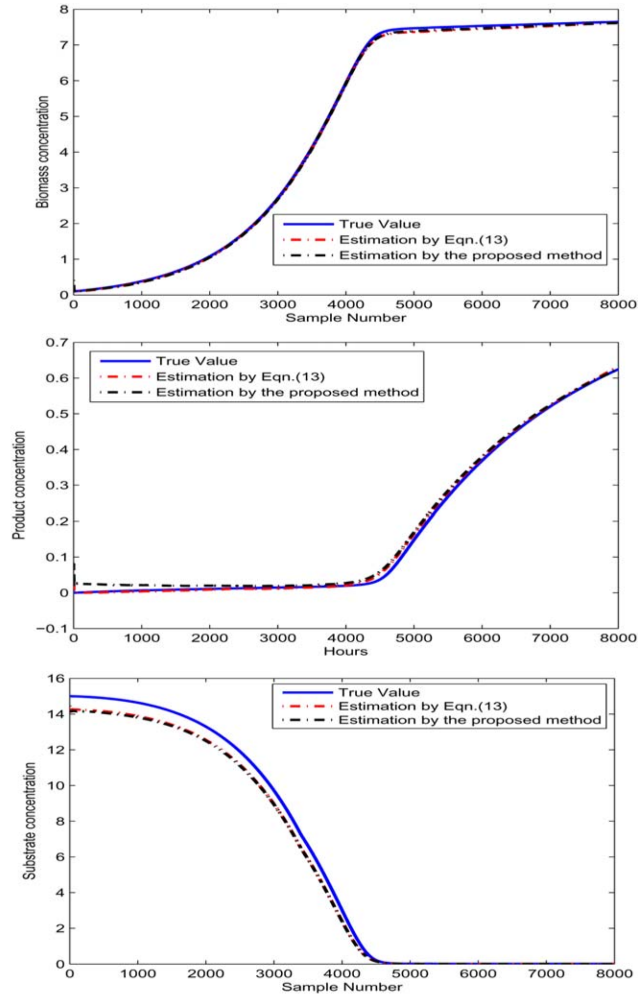

Figure 13. Estimation through imposing constraint on posterior particles, for example, 2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Conclusions

Three constrained strategies are proposed to impose constraints on the prior particles, the posterior particles and the state estimation, respectively. If the state is univariate or multivariate with no coupling constraint, constraints on the state are converted to those on the system noise, and then the constrained inverse transform sampling method is

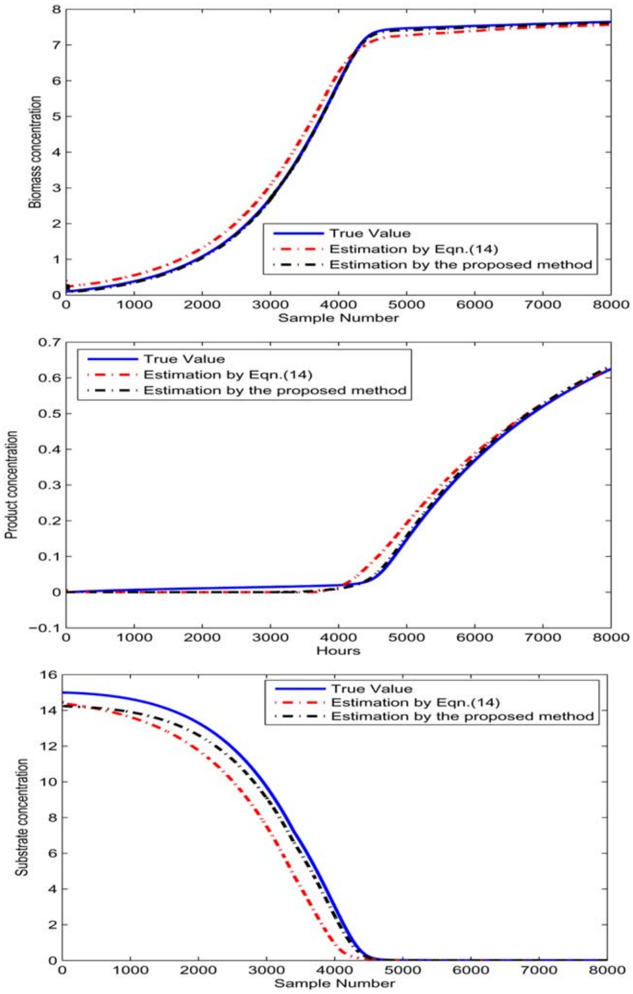

Figure 14. Estimation through imposing constraint on estimation, for example, 2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

proposed to produce prior particles. If the state is multivariate with coupling constraints, the Gibbs sampling method combined with the constrained inverse transform sampling method is proposed: the former transforms the generation of multivariate particles into that of univariate particles, and then the latter is used for the generation of valid prior particles. The constrained resampling method is proposed to restrict the resampling among valid prior particles to obtain valid posterior particles, which results in the similar form as the acceptance/rejection method in literature. Constraints are also imposed on state estimation by scaling the weights of valid posterior particles.

The proposed first scheme to constrain prior particles is suitable for nonlinear or linear inequality constraint. It extracts valid prior particles with better physical interpretation than the optimization-based method. However, as the generation of every particle involves the calculation of constraint region and inverse transformation, it will take more calculation time. The proposed second scheme has the same advantages and disadvantages as the acceptance/rejection method in literature. It is sensitive to the initial guess. If there are very limited valid initial prior particles, the quality of the posterior particles generated by the strategy will be poor. The proposed third scheme directly imposing constraint on the state estimation has no apparent limitation and can be applied to solve most types of constraint estimation problems, such as boundary constraint, nonlinear, and inequality constraints. As no assumption of Gaussian distribution is made and only the weights are adjusted rather than regenerating posterior particles, this proposed method provides three alterative schemes to implement the constrained state estimation with a better physic interpretation, no poorer performances than the existing method and no distributional assumptions on the particles.

Acknowledgments

The authors acknowledge the support in part by Natural Science and Engineering Research Council of Canada, and the 111 Project (B12018) as well as the support of Chang Jiang Scholar Program. The first author acknowledges the supports by Natural Science Foundation of Jiangsu Province (No. BK2010149) and National Natural Science Foundation of China (No. 60904045, NSFC 61134007).

Literature Cited

1. Doucet A, Freitas ND, Gordon N. *An Introduction to Sequential Monte Carlo Methods*. New York: Springer, 2001.
2. Ho YC, Lee RCK. A Bayesian approach to problems in stochastic estimation and control. *IEEE Trans Automat Control*. 1964;9(5): 333–339.
3. Chen Z. Bayesian filtering: from Kalman filter to particle filters, and beyond. Technical Report, Adaptive Systems Lab., McMaster University, 2003.
4. Gordon N, Salmond D, Smith AFM. Novel approach to nonlinear non Gaussian Bayesian state estimation. *IEE Proc F*. 1993;140(2): 107–113.
5. Barker AL, Brown DE, Martin WN. Bayesian estimation and the kalman filter. *Comput Math Appl*. 1995;30(10):55–77.
6. Jazwinski AH. *Stochastic Processes and Filtering Theory*. New York: Academic Press, 1970.
7. Kandepu R, Foss B, Imsland L. Applying the unscented Kalman filter for nonlinear state estimation. *J Process Control*. 2008;18: 753–768.
8. Qu C, Hahn J. Computation of arrival cost for moving horizon estimation via unscented Kalman filtering. *J Process Control*. 2009; 19(2):358–363.
9. Moshood JO. *Moving Horizon Estimation for Continuum and Noncontinuum States with Applications in Distillation Processes*. Ph.D. Thesis, University of Alberta, 2011.
10. Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking. *IEEE Trans Signal Process*. 2002;50(2):174–188.
11. Kitagawa G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J Comput Graph Stat*. 1996;5(1):1–25.
12. Chen WS, Bakshi BR. Bayesian estimation via sequential Monte Carlo sampling: unconstrained nonlinear dynamic systems. *Ind Eng Chem Res*. 2004;43:4012–4025.
13. Kandepu R, Imsland L, Foss B. Constrained state estimation using the unscented Kalman filter. In: *16th Mediterranean Conference on Control and Automation*, Ajaccio, France, 2008:1453–1458.
14. Kolas S, Foss B, Schei T. Constrained nonlinear state estimation based on the UKF approach. *Comput Chem Eng*. 2009;33(8):1386–1401.
15. Christopher VR, James BR, David QM. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Trans Automat Control*. 2003;48(2):246–258.
16. Ungarala S, Li K, Chen Z. Constrained Bayesian state estimation using a cell filter. *Ind Eng Chem Res*. 2008;47:7312–7322.
17. Lang L, Chen W, Bakshi B, Goel P, Ungarala S. Bayesian estimation via sequential Monte Carlo sampling-constrained dynamic systems. *Automatica*. 2007;43:1615–1622.
18. Shao X, Huang B, Lee JM. Constrained Bayesian state estimation—a comparative study and a new particle filter based approach. *J Process Control*. 2010;20:143–157.
19. Chen WS. Tutorial on sequential Monte Carlo sampling. Technical Report, Department of Chemical and Biomolecular Engineering, Ohio State University, 2002.

20. Kong A, Liu JS, Wong WH. Sequential imputations and Bayesian missing data problems. *J Am Stat Assoc.* 1994;89:278–288.
21. Doucet A, Godsill S. On sequential simulation-based methods for Bayesian filtering. Technical Report, Department of Engineering, Cambridge University, 1998.
22. Liu JS, Chen R. Sequential Monte Carlo methods for dynamical systems. *J Am Stat Assoc*. 1998;93:1032–1044.
23. Boloc M, Djuric P, Hong S. Resampling algorithms for particle filters: a computational complexity perspective. *EURASIP J Appl Signal Process.* 2004;15:2267–2277.
24. Smith AFM, Gelfand AE. Bayesian statistics without tears: a sampling-resampling perspective. *J Am Stat Assoc*. 1992;46(4):84–88.
25. Vachhani TP, Narasimhan S, Rengaswamy R. Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation. *J Process Control.* 2006;16:1075–1086.
26. Bϕlviken E, Acklam PJ, Christophersen N, Stϕrdal JM. Monte Carlo filters for non-linear state estimation. *Automatica.* 2001;37:177–183.
27. Chen T, Morris J, Martin E. Particle filters for state and parameter estimation in batch processes. *J Process Control* 2005;15:665–673.
28. Andrieu C, Doucet A, Singh S, Tadic VB. Particle methods for change detection, system identification, and control. *Proc IEEE*. 2004;92:423–437.
29. Tanizaki H. Nonlinear and non-Gaussian state space modeling using sampling techniques. *Ann Inst Stat Math.* 2001;53(1):63–81.
30. Sahu SK, Roberts GO. On convergence of the EM algorithm and the Gibbs sampler. *Stat Comput*. 1999;9(1):55–64.
31. Ritter C, Tanner MA. Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy-Gibbs sampler. *J Am Stat Assoc*. 1992;87(419):861–868.
32. Birol G, Undey C, Cinar A. A modular simulation package for fed-batch fermentation: penicillin production. *Comput Chem Eng*. 2000;26:1553–1565.
33. Zhao Z, Shao X, Huang B, Liu F. On-line estimation of glucose and biomass concentration in batch fermentation process using particle filter with constraint. *Asia Pac J Chem Eng*. 2012;7(5):678–686.

Manuscript received Oct. 8, 2013, and revision received Jan. 10, 2014.