

Sequential Monte Carlo Sampling in Hidden Markov Models of Nonlinear Dynamical Systems[☆]

X. Zeng^{a,b}, M. Anitescu^{b,*}

^a*Mathematics Department, Shanghai University, Shanghai, P.R. China*

^b*Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Building 240, Argonne, IL 60439-4844, U.S.A.*

Abstract

We investigate the issue of which state functionals can have their uncertainty estimated efficiently in dynamical systems with uncertainty. Because of the high dimensionality and complexity of the problem, sequential Monte Carlo (SMC) methods are used. We investigate SMC methods where the proposal distribution is computed by maximum likelihood or by a linearization approach. We prove that the variance of the SMC method is bounded linearly in the number of time steps when the proposal distribution is truncated normal distribution. We also show that for a moderate large number of steps the error produced by approximation of dynamical systems linearly accumulates on the condition that the logarithm of the density function of noise is Lipschitz continuous. This finding is significant because the uncertainty in many dynamical systems, in particular, in chemical engineering systems, can be assumed to have this nature. We demonstrate our findings for a simple test case from chemical engineering. The theoretical findings provide a foundation for the parallel software SISTOS.

Keywords: Hidden Markov model, ordinary differential equation, sequential Monte Carlo methods, chemical process

1. Introduction

Accountancy of expensive or consequential materials in the production process for chemical plants is a critical endeavor. For example, nuclear fuel reprocessing [1] facilities are expensive to build, are costly to operate, and have multiple stages and subsystems [2, 3, 4], losses of even minute amounts of the nuclear fuel can be therefore consequential. Currently, the United States has more than 15,000 chemical plant sites, which are required to file a risk management plan with the U.S. Environmental Protection Agency. Such plans consider both worst-case scenarios and alternative-case scenarios. Alternatives include the abnormal release of controlled materials or illegal plant interference.

[☆]Preprint ANL/MCS-P1852-0311, Argonne National Laboratory, Mathematics and Computer Science Division

*Corresponding author.

Email address: `anitescu@mcs.anl.gov` (M. Anitescu)

Preprint submitted to Elsevier

February 4, 2014

Whereas worst-case scenarios can be determined relatively efficiently with mass-balance approaches, it is difficult to detect long-term, slow releases, which are likely to be confused with measurement noise or other uncertain information. Hence, having superior accountancy will enable early detection of situations that may affect the safety of such plants and that may otherwise be hard to detect.

Aiming for incorporating more information in the accountancy process, we investigate issues connected with modeling chemical reactions, in addition to achieving a superior estimation of the state. *The problem is the one of estimating functionals of the state of a nonlinear dynamical system, given measurements with noise.* To that end, we investigate the use of a hidden Markov model to estimate important functionals of the state of the system [5]. However, because of the high dimensionality and the nonlinear nature of the chemical reaction, to produce samples from the target distribution is difficult. Not the least of the difficulties is the fact that the large dimension of high-fidelity models may not allow us to store in memory all estimates at one time. Therefore, we investigate sequential sampling plans. Specifically, sequential Monte Carlo (SMC) methods (also called particle filters) produce scenarios of past and current time so that function of interest can be estimated. The most famous filter is the Kalman filter [6], which is used to make estimates in a Bayesian framework for linear Gaussian problems. For nonlinear systems, ensemble Kalman filters [7, 8] or unscented Kalman filters [9] have emerged. One difficulty of such filters is their need to maintain at each step an approximation of the full covariance matrix and to update each of its entries, which results in exceedingly large memory and computational requirements [10].

In this paper, we investigate SMC methods that account for full nonlinearity. The proposal density is obtained by approximating the Hessian information of the logarithm of the targeted density at modal scores and can be loosely thought of as using a Kalman concept as a “preconditioner” for the sampling. By doing so, we get more accurate estimates while approaching the performance of Kalman methods. Our main goal is to determine whether SMC works for dynamical systems of the type encountered in chemical plants. We will thus justify the various assumptions we make about the dynamical systems by typical features encountered in real systems.

The key hypotheses investigated are that (1) the variance of the estimates produced by the method guarantees a priori the approach is tractable and (2) the proposal distributions we derive are sufficiently practical to be implemented. We demonstrate our findings on a synthetic example around the methane steam reforming reaction. We implement SISTOS, the Sequential Importance Sampling Toolkit of ODE Systems for statistical estimates of dynamical systems.

2. The Statistical/Uncertainty Model

2.1. Hidden Markov Model

The hidden Markov model (HMM) is a widely used approach in temporal pattern recognition, such as finance [11]; handwriting [12]; speech [13]; bioinformatics [14]; and, perhaps the area closest to that used here, weather forecast [10].

HMM is based on a state space model that includes two components. The first describes the state evolution of the dynamical system; while the second models noisy

observations that are dependent on the state. Both parts include uncertainties. The mathematical formulation is described as follows

$$\mathbf{x}_i = M(\mathbf{x}_{i-1}) + \boldsymbol{\mu}_i, \quad (1)$$

$$\mathbf{z}_i = H(\mathbf{x}_i) + \boldsymbol{\nu}_i, \quad (2)$$

where i is the discrete time index. Here $\mathbf{x}_i \in \mathcal{X}$ is the state variable (for example, the moles of the reactants in the reaction at the beginning of step i) and $\boldsymbol{\mu}_i$ is a random variable that quantifies the uncertainty in the model (for example, in the outside feed or in the model itself). The variable $\mathbf{z}_i \in \mathcal{Z}$ are the observed quantities (for example, the outputs after some time period t_f) and $\boldsymbol{\nu}_i$ is a random variable that quantifies measurement error i . *The variables $\boldsymbol{\mu}_i$ and $\boldsymbol{\nu}_i$ are the core of the uncertainty model.* Their definition controls the lack of information about both the computational model and the measurement process, essentially the entire uncertainty space.

Another component of the model is $M : \mathcal{M} \subseteq \mathcal{X} \rightarrow \mathcal{X}$, a nonlinear function describing the evolution of the dynamical system over the period between two observations (for simplicity we assume it models an autonomous system, though that assumption can easily be relaxed). The mapping $H : \mathcal{X} \rightarrow \mathcal{Z}$ denotes the response of the measuring process. It is natural to assume that the $\mathcal{X} \subset \mathcal{R}^{n_x}$ and $\mathcal{Z} \subset \mathcal{R}^{n_z}$ are compact because many dynamical systems such as chemical plants have limited capacity and most state variables should be nonnegative.

It immediately follows that the system described in (1) has the Markov property

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \dots, \mathbf{x}_0) = p(\mathbf{x}_k | \mathbf{x}_{k-1}).$$

To simplify the notation, we let $\mathbf{x}_{0:k} := (\mathbf{x}_0, \dots, \mathbf{x}_k)^T$ and $\mathbf{z}_{0:k} := (\mathbf{z}_0, \dots, \mathbf{z}_k)^T$. From the Markov property we obtain that the prior density is

$$p(\mathbf{x}_{0:k}) = \prod_{i=0}^k p(\mathbf{x}_i | \mathbf{x}_{i-1}). \quad (3)$$

From (2) we compute the density of the collection of observations conditioned on $\mathbf{x}_{0:k}$ as

$$p(\mathbf{z}_{0:k} | \mathbf{x}_{0:k}) = \prod_{i=0}^k p(\mathbf{z}_i | \mathbf{x}_i). \quad (4)$$

Here the density function $p(\mathbf{x}_i | \mathbf{x}_{i-1})$ and $p(\mathbf{z}_i | \mathbf{x}_i)$ are calculated from the distribution of $\boldsymbol{\mu}_i$ and $\boldsymbol{\nu}_i$ by using the convention $p(\mathbf{x}_0 | \mathbf{x}_{-1}) = p(\mathbf{x}_0)$.

According to the Bayesian formula, we compute the posterior predictor density or targeted quantity

$$p(\mathbf{x}_{0:k} | \mathbf{z}_{0:k}) = \frac{p(\mathbf{x}_{0:k})p(\mathbf{z}_{0:k} | \mathbf{x}_{0:k})}{p(\mathbf{z}_{0:k})}. \quad (5)$$

The numerator of (5) can be readily calculated through equations (3) and (4). The denominator is difficult to obtain, however, since it requires an integration over a large dimensional space with density $p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k})$.

If both the error model and the noise are normally distributed, that is, $\boldsymbol{\mu}_i \sim \mathcal{N}(\bar{\boldsymbol{x}}_i, Q_i)$ and $\boldsymbol{\nu}_i \sim \mathcal{N}(0, R_i)$, the posterior density function then can be written as follows:

$$p(\boldsymbol{x}_{0:k} | \boldsymbol{z}_{0:k}) \propto \frac{\exp\left(-\frac{1}{2} \sum_{i=0}^k g_i(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i, \boldsymbol{z}_i)\right)}{p(\boldsymbol{z}_{0:k})}, \quad (6)$$

where

$$g_i(\boldsymbol{x}_{i-1}, \boldsymbol{x}_i, \boldsymbol{z}_i) = (\boldsymbol{x}_i - M(\boldsymbol{x}_{i-1}) - \bar{\boldsymbol{x}}_i)^T Q_i^{-1} (\boldsymbol{x}_i - M(\boldsymbol{x}_{i-1}) - \bar{\boldsymbol{x}}_i) + (\boldsymbol{z}_i - H(\boldsymbol{x}_i))^T R_i^{-1} (\boldsymbol{z}_i - H(\boldsymbol{x}_i)) \quad (7)$$

with $M(\boldsymbol{x}_{-1}) = 0$.

2.2. The Problem to Be Solved

The formal problem is stated as follows: Evaluate the integral

$$I(\phi, k, p) = \int \phi(\boldsymbol{x}_{0:k}) p(\boldsymbol{x}_{0:k} | \boldsymbol{z}_{0:k}) d\boldsymbol{x}_{0:k} \quad (8)$$

given the hidden Markov model and a sequence of observations $\boldsymbol{z}_{0:k}$. In our work, two types of ϕ are considered. One is a cumulative state estimate,

$$\phi(\boldsymbol{x}_{0:k}) = \sum_{i=0}^k \psi(\boldsymbol{x}_i); \quad (9)$$

the other is terminal state estimate

$$\phi(\boldsymbol{x}_{0:k}) = \psi(\boldsymbol{x}_k). \quad (10)$$

3. The Sequential Monte Carlo Method

The early idea of the sequential Monte Carlo methodology is provided in [15] and [16]. Since then, these kinds of methods have received considerable attention and are widely used in automatic control [17], geophysical group [18], and biology [19]. Sequential Monte Carlo (SMC) methods are sophisticated estimation techniques. Such methods are usually employed to produce samples (particles) of state variables or their associated calculated values for hidden Markov models. These samples are propagated over time by using the importance sampling (IS) [20] and resampling mechanisms [21].

According to [22], the following recursive formula holds:

$$p(\boldsymbol{x}_{0:i} | \boldsymbol{z}_{0:i}) = p(\boldsymbol{x}_{0:i-1} | \boldsymbol{z}_{0:i-1}) \frac{p(\boldsymbol{z}_i | \boldsymbol{x}_i) p(\boldsymbol{x}_i | \boldsymbol{x}_{i-1})}{p(\boldsymbol{z}_i | \boldsymbol{z}_{0:i-1})}, \quad (11)$$

If we can produce samples at time step i from

$$\frac{p(\boldsymbol{z}_i | \boldsymbol{x}_i) p(\boldsymbol{x}_i | \boldsymbol{x}_{i-1})}{p(\boldsymbol{z}_i | \boldsymbol{z}_{1:i-1})}, \quad (12)$$

then (11) can be used to recursively sample from the proper distribution. Nevertheless, the denominator in (12) is exceedingly hard to estimate, thus making this distribution difficult to sample from. This difficulty is mitigated by importance sampling.

The basic idea of importance sampling, which is used here, is to sample from a proposal distribution so that its samples make the integrand large and make the most important contribution to the integrand (8). Suppose that at the i th step, with known states for previous steps and the observations until now, the proposal density is denoted as $q(\mathbf{x}_i|\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i})$. The proposal density chosen here is a general case. Later we will choose a particular case that is much easier to compute. Define

$$q(\mathbf{x}_{0:i}|\mathbf{z}_{0:i}) := q(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1})q(\mathbf{x}_i|\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i})$$

to be the proposal density for $p(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})$. The ratio of $p(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})$ and $q(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})$ can be computed recursively as

$$\frac{p(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})}{q(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})} = \frac{p(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1})}{q(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1})} \frac{p(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})}{p(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1})q(\mathbf{x}_i|\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i})}.$$

By rearranging this equation, we obtain

$$\frac{p(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})}{q(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})} = \frac{p(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1})}{q(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1})} \frac{p(\mathbf{x}_{0:i}, \mathbf{z}_{0:i})}{p(\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i-1})q(\mathbf{x}_i|\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i})} \frac{1}{p(\mathbf{z}_i|\mathbf{z}_{0:i-1})}.$$

This formula suggests a sampling-based approximation of the conditioned density in terms of the sequential sampling as follows:

$$p(\mathbf{x}_{0:i}|\mathbf{z}_{0:i}) = \sum_{n=1}^N W_i^n \delta_{\mathbf{x}_{0:i}^n}(\mathbf{x}_{0:i}),$$

where $\mathbf{x}_{0:i}^n$ are sampled from proposal density q . Of course $p(\mathbf{z}_i|\mathbf{z}_{0:i-1})$ is unknown, but it is only a number that can be eliminated by normalizing the weight to add to 1. Thus, the unnormalized weight of the n th sample, $\mathbf{x}_{0:i}^n$, is defined by

$$\widetilde{W}_i^n = W_{i-1}^n \frac{p(\mathbf{x}_{0:i}^n, \mathbf{z}_{0:i})}{p(\mathbf{x}_{0:i-1}^n, \mathbf{z}_{0:i-1})q(\mathbf{x}_i^n|\mathbf{x}_{0:i-1}^n, \mathbf{z}_{0:i})}, \quad (13)$$

and the normalized weight is obtained by

$$W_i^n = \frac{\widetilde{W}_i^n}{\sum_{n=1}^N \widetilde{W}_i^n}. \quad (14)$$

Note that (13) involves only unnormalized weights and hence uses only joint density functions $p(\mathbf{x}_{0:i}^n, \mathbf{z}_{0:i})$ and $p(\mathbf{x}_{0:i-1}^n, \mathbf{z}_{0:i-1})$ and is easy to compute. With the samples sequentially produced from proposal density and the weights, we can compute a Monte Carlo approximation of the integral:

$$I(\phi, k, p) \approx \sum_{n=1}^N W_k^n \phi(\mathbf{x}_{0:k}^n). \quad (15)$$

This method is called sequential importance sampling (SIS).

SIS usually is good only for moderate-size problems because the variance of the weights tends to increase when k increases. Resampling would improve the performance by removing samples with possible low weight. The basic idea of resampling is to use the information of weights of samples of previous steps to get the new samples up to current time step so that every new sample has a new equal weight. The most commonly used resampling methods are systematic random resampling, residual resampling and multinomial resampling [21, 22]. In our context, we call SIS methods plus resampling SMC methods [22]. After resampling, the weights are updated to be equal, that is, $1/N$. Then the approximation to the integral is

$$I^{\text{smc}} = \sum_{n=1}^N \frac{1}{N} \phi(\mathbf{x}_{0:k}^n). \quad (16)$$

3.1. Generating the Proposal Distribution by Maximum Likelihood

In this subsection, our goal is to develop a method to find a proposal density by obtaining a second-order approximation of $\log [p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i)]$. To simplify the problem, we assume that the following regularity conditions are satisfied by $\log [p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i)]$.

Assumption 1. *The log density $\log [p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i)]$ is twice continuously differentiable with respect to \mathbf{x}_i . There exists a unique solution of $\nabla_{\mathbf{x}_i} \log (p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i)) = 0$ and its Hessian $\nabla_{\mathbf{x}_i}^2 \log (p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i))$ at the solution is negative definite.*

We choose a special proposal density $q(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{z}_i)$ by using the maximum likelihood method. The proposal density $q(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{z}_i)$ is taken as the density function of normal distribution $\mathcal{N}(\mathbf{x}_i^*, \Theta_i(\mathbf{x}_i^*))$, where

$$\begin{aligned} \mathbf{x}_i^* &= \operatorname{argmax}_{\mathbf{x}_i} \{p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i)\}, \\ \Theta_i^{-1}(\mathbf{x}_i^*) &:= \nabla_{\mathbf{x}_i}^2 \log (p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i))|_{\mathbf{x}_i=\mathbf{x}_i^*}. \end{aligned}$$

We note that this approximation is much easier to calculate than the density conditional on all observations. We take the normal noise and model error example (6)–(7) as a demonstration. We have at the i th time step

$$p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i) \propto \exp\left(-\frac{1}{2}g_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i)\right),$$

where g_i is defined in (7). The gradient of $\frac{1}{2}g_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i)$ with respect to \mathbf{x}_i is

$$\begin{aligned} \nabla_{\mathbf{x}_i} \left(\frac{1}{2}g_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i)\right) &= (\mathbf{x}_i - M(\mathbf{x}_{i-1}) - \bar{\mathbf{x}}_i)^T Q_i^{-1} + \\ &\quad (H(\mathbf{x}_i) - \mathbf{z}_i)^T R_i^{-1} \nabla_{\mathbf{x}_i} H(\mathbf{x}_i), \end{aligned} \quad (17)$$

and the Hessian of $\frac{1}{2}g_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i)$ with respect to \mathbf{x}_i is

$$\begin{aligned} \nabla_{\mathbf{x}_i}^2 \left(\frac{1}{2}g_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i)\right) &= Q_i^{-1} + (\nabla_{\mathbf{x}_i} H(\mathbf{x}_i))^T R_i^{-1} \nabla_{\mathbf{x}_i} H(\mathbf{x}_i) + \\ &\quad ((R_i^{-1}(H(\mathbf{x}_i) - \mathbf{z}_i))^T \otimes I_{n_x}) \nabla_{\mathbf{x}_i}^2 H. \end{aligned} \quad (18)$$

We can easily verify that Assumption 1 is satisfied if the Hessian matrix (18) is positive definite. We can find the minimizer by setting its gradient (17) to be zero. Its solution is denoted as \mathbf{x}_i^* . Applying the Taylor expansion to g_i at the point \mathbf{x}_i^* , we then have

$$g_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i) \approx g_i(\mathbf{x}_{i-1}, \mathbf{x}_i^*, \mathbf{z}_i) + (\mathbf{x}_i - \mathbf{x}_i^*)^T \Theta_i^{-1}(\mathbf{x}_i^*)(\mathbf{x}_i - \mathbf{x}_i^*),$$

where $\Theta_i^{-1}(\mathbf{x}_i^*)$ is equal to the Hessian matrix of $\frac{1}{2}g_i$ at \mathbf{x}_i^* , that is,

$$\Theta_i^{-1}(\mathbf{x}_i^*) := \nabla_{\mathbf{x}_i}^2 \left(\frac{1}{2}g_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i) \right) \Big|_{\mathbf{x}_i=\mathbf{x}_i^*}. \quad (19)$$

We then can use $\mathcal{N}(\mathbf{x}_i^*, \Theta_i(\mathbf{x}_i^*))$ as the proposal distribution at the i th step.

3.2. Generating the Proposal Distribution by a Linearization Method

If $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are normally distributed and the mean of $\boldsymbol{\mu}_i$ is $\bar{\mathbf{x}}_i$, according to (7), we can get a quadratic polynomial of \mathbf{x}_i by Taylor expansion of $H(\mathbf{x}_i)$ at $M(\mathbf{x}_{i-1}) + \bar{\mathbf{x}}_i$ as

$$H(\mathbf{x}_i) \approx \hat{H}(\mathbf{x}_i) := H(M(\mathbf{x}_{i-1}) + \bar{\mathbf{x}}_i) + A_i(\mathbf{x}_i - M(\mathbf{x}_{i-1}) - \bar{\mathbf{x}}_i),$$

where $A_i = \nabla H(M(\mathbf{x}_{i-1}) + \bar{\mathbf{x}}_i)$. We therefore can obtain the following quadratic as an approximation of g_i in (7):

$$\begin{aligned} \hat{g}_i(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_i) &:= (\mathbf{x}_i - M(\mathbf{x}_{i-1}) - \bar{\mathbf{x}}_i)^T Q_i^{-1} (\mathbf{x}_i - M(\mathbf{x}_{i-1}) - \bar{\mathbf{x}}_i) \\ &\quad + (\mathbf{z}_i - \hat{H}(\mathbf{x}_i))^T R_i^{-1} (\mathbf{z}_i - \hat{H}(\mathbf{x}_i)). \end{aligned}$$

Thus, we can use the density function of $\mathcal{N}(\hat{\mathbf{x}}_i, \Sigma_i)$ as proposal density $q(\mathbf{x}_i | \mathbf{x}_{i-1}, \mathbf{z}_{0:i})$, where

$$\hat{\mathbf{x}}_i = M(\mathbf{x}_{i-1}) + \bar{\mathbf{x}}_i + (Q_i^{-1} + A_i^T R_i^{-1} A_i)^{-1} A_i^T R_i^{-1} (\mathbf{z}_i - H(M(\mathbf{x}_{i-1}) + \bar{\mathbf{x}}_i)), \quad (20)$$

and

$$\Sigma_i^{-1} := Q_i^{-1} + A_i^T R_i^{-1} A_i. \quad (21)$$

The two methods introduced in Sections 3.1 and 3.2 use the information of derivatives of function H or M . In applications, $M(\mathbf{x})$, $H(\mathbf{x})$ can involve ordinary differential equations. In this case, in order to compute the derivatives, the first or second sensitivities [23] are required.

3.3. Truncated Normal Distribution

For importance sampling methods to work efficiently, one needs to make sure that the denominators are far from zero when computing the weights (13). One way to ensure such an occurrence is by having a *heavy-tailed* proposal distribution: one whose tails are heavier than or comparable to the ones of the target distribution. Another way is to focus on problems with bounded noise as long as the proposal distribution has density bounded away from zero on the compact domain of the noise. For the chemical plant application covered in this paper, both $\boldsymbol{\mu}_i$, and $\boldsymbol{\nu}_i$ are naturally bounded, therefore the latter case does apply. One way to obtain distributions that have densities bounded below in this compact domain is to use truncated normal distributions, which is what we will

do for numerical studies in this paper. As domains, we choose $\mathcal{D}_{\boldsymbol{\mu}_i} := \{0 \leq \boldsymbol{\mu}_i \leq \bar{\boldsymbol{x}}_i + \boldsymbol{\zeta}_i\}$ for $\boldsymbol{\mu}_i$ since $\boldsymbol{\mu}_i$ represents material quantities (should greater than zero) in applications, and $\mathcal{D}_{\boldsymbol{\nu}_i} := \{-\boldsymbol{\delta}'_i \leq \boldsymbol{\nu}_i \leq \boldsymbol{\delta}_i\}$ for some positive vector $\boldsymbol{\delta}_i$ for $\boldsymbol{\nu}_i$ that denotes measurement noise. We then have the truncated normal density:

$$p_{\boldsymbol{\mu}_i}(\boldsymbol{\mu}_i) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T \boldsymbol{\Theta}_{\boldsymbol{\mu}_i}^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right) \delta_{\mathcal{D}_{\boldsymbol{\mu}_i}}}{\int_{\mathcal{D}_{\boldsymbol{\mu}_i}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_i)^T \boldsymbol{\Theta}_{\boldsymbol{\mu}_i}^{-1}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_i)\right) d\boldsymbol{\mu}} \quad (22)$$

and

$$p_{\boldsymbol{\nu}_i}(\boldsymbol{\nu}_i) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{\nu}_i - \bar{\boldsymbol{\nu}}_i)^T \boldsymbol{\Theta}_{\boldsymbol{\nu}_i}^{-1}(\boldsymbol{\nu}_i - \bar{\boldsymbol{\nu}}_i)\right) \delta_{\mathcal{D}_{\boldsymbol{\nu}_i}}}{\int_{\mathcal{D}_{\boldsymbol{\nu}_i}} \exp\left(-\frac{1}{2}(\boldsymbol{\nu} - \bar{\boldsymbol{\nu}}_i)^T \boldsymbol{\Theta}_{\boldsymbol{\nu}_i}^{-1}(\boldsymbol{\nu} - \bar{\boldsymbol{\nu}}_i)\right) d\boldsymbol{\nu}}. \quad (23)$$

Let $\mathcal{D}_i = \{\boldsymbol{x}_i - M(\boldsymbol{x}_{i-1}) \in \mathcal{D}_{\boldsymbol{\mu}_i}\} \cap \{\boldsymbol{z}_i - H(\boldsymbol{x}_i) \in \mathcal{D}_{\boldsymbol{\nu}_i}\}$. It can be proved by induction that \mathcal{D}_i is still a bounded domain if $M(\cdot)$ is bounded and $\mathcal{D}_{\boldsymbol{\mu}_i}, \mathcal{D}_{\boldsymbol{\nu}_i}$ are bounded. The proposal density takes the form

$$q(\boldsymbol{x}_i | \boldsymbol{x}_{i-1}, \boldsymbol{z}_i) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{m}_i)^T \boldsymbol{\Theta}_i^{-1}(\boldsymbol{x}_i - \boldsymbol{m}_i)\right) \delta_{\mathcal{D}_i}}{\int_{\mathcal{D}_i} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m}_i)^T \boldsymbol{\Theta}_i^{-1}(\boldsymbol{x} - \boldsymbol{m}_i)\right) d\boldsymbol{x}}. \quad (24)$$

Here \boldsymbol{m}_i and $\boldsymbol{\Theta}_i^{-1}$ are determined by the maximum likelihood method or linearization method.

3.4. Convergence Results of the SMC Method

Convergence results for the sequential Monte Carlo method are discussed in [22], [24], [25] and [26]. We now adapt them to the case discussed in this paper. The general convergence formulae using the multinomial resampling are given in [22] and [24]. That is, if the multinomial resampling method is used in selection, then

$$N^{1/2}(I^{\text{smc}}(\phi, k, p) - I(\phi, k, p))$$

converges to $\mathcal{N}(0, V_k)$ in distribution as $N \rightarrow +\infty$. Here, for $\phi : \boldsymbol{x}_{0:k} \rightarrow \phi(\boldsymbol{x}_{0:k})$, we define the following:

$$V_k = \int \frac{p^2(\boldsymbol{x}_{0:k} | \boldsymbol{z}_{0:k})}{q(\boldsymbol{x}_0 | \boldsymbol{z}_0)} \left(\int \phi(\boldsymbol{x}_{0:k}) p(\boldsymbol{x}_{1:k} | \boldsymbol{x}_0, \boldsymbol{z}_{0:k}) d\boldsymbol{x}_{1:k} - I(\phi, k, p) \right)^2 d\boldsymbol{x}_0 \\ + \sum_{i=1}^{k-1} \int \frac{p(\boldsymbol{x}_{0:i} | \boldsymbol{z}_{0:k})}{p(\boldsymbol{x}_{0:i-1} | \boldsymbol{z}_{0:i-1}) q(\boldsymbol{x}_i | \boldsymbol{x}_{0:i-1}, \boldsymbol{z}_{0:i})} \times \quad (25)$$

$$\left(\int \phi(\boldsymbol{x}_{0:k}) p(\boldsymbol{x}_{i+1:k} | \boldsymbol{x}_{0:i}, \boldsymbol{z}_{0:k}) d\boldsymbol{x}_{i+1:k} - I(\phi, k, p) \right)^2 p(\boldsymbol{x}_{0:i} | \boldsymbol{z}_{0:k}) d\boldsymbol{x}_{0:i} \\ + \int \frac{p^2(\boldsymbol{x}_{0:k} | \boldsymbol{z}_{0:k})}{p(\boldsymbol{x}_{0:k-1} | \boldsymbol{z}_{0:k-1}) q(\boldsymbol{x}_k | \boldsymbol{x}_{0:k-1}, \boldsymbol{z}_{0:k})} (\phi(\boldsymbol{x}_{0:k}) - I(\phi, k, p))^2 d\boldsymbol{x}_{0:k}. \quad (26)$$

If $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_k)$, then

$$\begin{aligned}
V_k &= \int \frac{p^2(\mathbf{x}_0|\mathbf{z}_{0:k})}{q(\mathbf{x}_0|\mathbf{z}_0)} \left(\int \phi(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_0, \mathbf{z}_{0:k})d\mathbf{x}_k - I(\phi, k, p) \right)^2 d\mathbf{x}_0 \\
&\quad + \sum_{i=1}^{k-1} \int \frac{p(\mathbf{x}_{i-1:i}|\mathbf{z}_{0:k})}{p(\mathbf{x}_{i-1}|\mathbf{z}_{0:i-1})q(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{z}_{0:i})} \times \\
&\quad \quad \left(\int \phi(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_i, \mathbf{z}_{0:k})d\mathbf{x}_k - I(\phi, k, p) \right)^2 p(\mathbf{x}_{i-1:i}|\mathbf{z}_{0:k})d\mathbf{x}_{i-1:i} \\
&\quad + \int \frac{p^2(\mathbf{x}_{k-1:k}|\mathbf{z}_{0:k})}{p(\mathbf{x}_{k-1}|\mathbf{z}_{0:k-1})q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{0:k})} (\phi(\mathbf{x}_k) - I(\phi, k, p))^2 d\mathbf{x}_{k-1:k}. \tag{27}
\end{aligned}$$

Note that (25) is derived for general proposal densities; that is, proposal densities are used as a form of $q(\mathbf{x}_i|\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i})$ and (27) is derived for proposal densities that are formed as $q(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{z}_{0:i})$. Chopin showed that the asymptotic variance using residual resampling is smaller than that using multinomial resampling in [24]. Kunsch showed that residual resampling and systematic resampling improve convergence results in [26].

One can tell from (25) and (27) that V_k increases with k . According to [27] sequential Monte Carlo methods would fail in some cases because of V_k 's increases at exponential rate. Hence, to make sure that SMC is tractable, we want to investigate under what condition can V_k be assumed to not increase rapidly with k . We do so by imposing reasonable conditions on the distribution of noise and proposal distribution. For V_k defined in (25), we have the following lemma.

Lemma 1. *Assume that for*

$$p(\mathbf{z}_i|\mathbf{x}_i) \geq C_1 > 0, \text{ and } 0 < C_2 < p(\mathbf{x}_{i+1}|\mathbf{x}_i) < C_3, \quad i = 0, \dots, k, \tag{28}$$

and

$$\frac{p(\mathbf{x}_i|\mathbf{x}_{i-1})p(\mathbf{z}_i|\mathbf{x}_i)}{q(\mathbf{x}_i|\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i})} \leq C_4. \tag{29}$$

Then for $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_{0:k})$ and $\|\phi\| = \max \phi(\mathbf{x}_{0:k}) < \infty$, we have that

$$V_k \leq (C_1 C_2)^{-1} C_3 C_4 (k+1) \|\phi\|.$$

Therefore, the variance increases no faster than linearly with the number of time steps. Thus the complexity failure of the exponential type is avoided.

Proof. Let Z_i to denote the normalizing constant $p(\mathbf{z}_{0:i})$ of $p(\mathbf{x}_{0:i}|\mathbf{z}_{0:i})$, that is,

$$Z_i := p(\mathbf{z}_{0:i}) = \int p(\mathbf{x}_{0:i}, \mathbf{z}_{0:i})d\mathbf{x}_{0:i} = \int \prod_{j=0}^i p(\mathbf{x}_j|\mathbf{x}_{j-1})p(\mathbf{z}_j|\mathbf{x}_j)d\mathbf{x}_{0:i}. \tag{30}$$

We can find an upper bound of Z_k by using $p(\mathbf{z}_i|\mathbf{x}_i)p(\mathbf{x}_{i+1}|\mathbf{x}_i) > C_1 C_2$ and taking them out of the integrand. That is,

$$Z_k > C_1 C_2 Z_{i-1} \int p(\mathbf{z}_{i+1}|\mathbf{x}_{i+1}) \prod_{j=i+2}^k p(\mathbf{x}_j|\mathbf{x}_{j-1})p(\mathbf{z}_j|\mathbf{x}_j)d\mathbf{x}_{i+1:k}. \tag{31}$$

According to the definition of $p(\mathbf{x}_{0:i}|\mathbf{z}_{0:k})$, we have

$$\begin{aligned}
p(\mathbf{x}_{0:i}|\mathbf{z}_{0:k}) &= Z_k^{-1} \int p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{i+1:k} \\
&= Z_k^{-1} p(\mathbf{x}_{0:i}, \mathbf{z}_{0:i}) \int \prod_{j=i+1}^k p(\mathbf{x}_j|\mathbf{x}_{j-1}) p(\mathbf{z}_j|\mathbf{x}_j) d\mathbf{x}_{i+1:k} \\
&\leq Z_k^{-1} C_3 p(\mathbf{x}_{0:i}, \mathbf{z}_{0:i}) \int p(\mathbf{z}_{i+1}|\mathbf{x}_{i+1}) \prod_{j=i+2}^k p(\mathbf{x}_j|\mathbf{x}_{j-1}) p(\mathbf{z}_j|\mathbf{x}_j) d\mathbf{x}_{i+1:k} \\
&\leq Z_{i-1}^{-1} (C_1 C_2)^{-1} C_3 p(\mathbf{x}_{0:i}, \mathbf{z}_{0:i}). \tag{32}
\end{aligned}$$

The last equality holds from (31). According to Bayes' formula, we have

$$p(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1}) = Z_{i-1}^{-1} p(\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i-1}). \tag{33}$$

Hence, from (32), (33), and condition (29), we have

$$\frac{p(\mathbf{x}_{0:i}|\mathbf{z}_{0:k})}{p(\mathbf{x}_{0:i-1}|\mathbf{z}_{0:i-1}) q(\mathbf{x}_i|\mathbf{x}_{0:i-1}, \mathbf{z}_{0:i})} \leq (C_1 C_2)^{-1} C_3 C_4. \tag{34}$$

Because $\|\phi\| = \max \phi(\mathbf{x}_{0:k}) < \infty$, we have that

$$\left(\int \phi(\mathbf{x}_{0:k}) p(\mathbf{x}_{i+1:k}|\mathbf{x}_{0:i}, \mathbf{z}_{0:k}) d\mathbf{x}_{i+1:k} - I(\phi, k, p) \right)^2 \leq 4\|\phi\|^2. \tag{35}$$

The last inequality follows from the triangle inequality of absolute values and the fact that $|\int \psi(\mathbf{x}, \mathbf{y}) g(\mathbf{x}|\mathbf{z}) d\mathbf{x}| \leq C_u \int g(\mathbf{x}|\mathbf{z}) d\mathbf{x} = C_u$ if $\psi(\mathbf{x}, \mathbf{y}) \leq C_u$ and g is a density function, which is also applied to every component in (25). Thus, from (34) and (35), each component of variance V_k in (25) is upper bounded by $4(C_1 C_2)^{-1} C_3 C_4 \|\phi\|^2$, which implies that

$$V_k \leq 4(C_1 C_2)^{-1} C_3 C_4 (k+1) \|\phi\|^2.$$

The proof is then complete. \square

Condition (28) is not satisfied by the normal distribution. But it is satisfied for the truncated normal distribution in (22) and (23) if $H(\cdot)$ and $M(\cdot)$ are continuous and $\mathcal{D}_{\mu_i}, \mathcal{D}_{\nu_i}$ are bounded.

For V_k defined in (27), a better result can be achieved if the distribution of noise can “forget the past.” This ability is crucial in obtaining our estimate in Lemma 2. To measure this ability, we can use contraction coefficients [28]. According to [24], the contraction coefficients ρ_k satisfy

$$\rho_k := \frac{1}{2} \sup_{\mathbf{x}'_l, \mathbf{x}''_l \in \mathcal{X}_l} \int |p(\mathbf{x}_k|\mathbf{x}'_l, \mathbf{z}_{0:k}) - p(\mathbf{x}_k|\mathbf{x}''_l, \mathbf{z}_{0:k})| d\mathbf{x}_k \leq (1 - C^{-2})^{k-l}, \tag{36}$$

if the following inequality holds:

$$\frac{p(\mathbf{x}_i|\mathbf{x}'_{i-1})}{p(\mathbf{x}_i|\mathbf{x}''_{i-1})} \leq C, \forall \mathbf{x}_i, \mathbf{x}'_{i-1}, \mathbf{x}''_{i-1} \in \mathcal{X}_i, \quad \forall i = 1, \dots, k. \tag{37}$$

The slightly stronger condition

$$c_a h(\mathbf{x}_i) \leq p(\mathbf{x}_i | \mathbf{x}_{i-1}) \leq C_a h(\mathbf{x}_i), \quad (38)$$

is provided by Kunsch in [26] as an alternative to (37). Kunsch proved that condition (38) is satisfied for bounded M and $\boldsymbol{\mu}_i$ and when the logarithm of density function is uniformly Lipschitz continuous. Most heavy-tailed distributions and the truncated normal distribution discussed in the paper, satisfy this condition, but not normal distributions.

Lemma 2. *If (28) and (37) hold, and if*

$$\frac{p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{z}_i | \mathbf{x}_i)}{q(\mathbf{x}_i | \mathbf{x}_{i-1}, \mathbf{z}_{0:i})} \leq C_4, \quad (39)$$

then V_k defined in (27) for $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_k)$ with $\phi \leq \|\phi\|$ is bounded.

Proof. First, the difference between $\int \phi(\mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}_i, \mathbf{z}_{0:k}) d\mathbf{x}_k$ and $I(\phi, k, p)$ can be bounded by

$$\begin{aligned} & \left| \int \phi(\mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}_i, \mathbf{z}_{0:k}) d\mathbf{x}_k - I(\phi, k, p) \right| \\ &= \left| \int \phi(\mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}_i, \mathbf{z}_{0:k}) d\mathbf{x}_k - \int \phi(\mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}'_i, \mathbf{z}_{0:k}) p(\mathbf{x}'_i | \mathbf{z}_{0:k}) d\mathbf{x}'_i d\mathbf{x}_k \right| \\ &= \left| \int \phi(\mathbf{x}_k) \left(p(\mathbf{x}_k | \mathbf{x}_i, \mathbf{z}_{0:k}) - \int p(\mathbf{x}_k | \mathbf{x}'_i, \mathbf{z}_{0:k}) p(\mathbf{x}'_i | \mathbf{z}_{0:k}) d\mathbf{x}'_i \right) d\mathbf{x}_k \right| \\ &\leq \|\phi\| \int |p(\mathbf{x}_k | \mathbf{x}_i, \mathbf{z}_{0:k}) p(\mathbf{x}'_i | \mathbf{z}_{0:k}) - p(\mathbf{x}_k | \mathbf{x}'_i, \mathbf{z}_{0:k}) p(\mathbf{x}'_i | \mathbf{z}_{0:k})| d\mathbf{x}'_i d\mathbf{x}_k \\ &= \|\phi\| \int p(\mathbf{x}'_i | \mathbf{z}_{0:k}) d\mathbf{x}'_i \int |p(\mathbf{x}_k | \mathbf{x}_i, \mathbf{z}_{0:k}) - p(\mathbf{x}_k | \mathbf{x}'_i, \mathbf{z}_{0:k})| d\mathbf{x}_k \\ &\leq 2\|\phi\|(1 - C^{-2})^{k-i}. \end{aligned} \quad (40)$$

Note that this inequality holds for any \mathbf{x}_i . The last relationship follows from (36) and from the cited result of [24].

By using the definition of $p(\mathbf{x}_{i-1:i} | \mathbf{z}_{0:k})$ and a technique similar to that used in the proof of Lemma 1, we have that

$$\begin{aligned} & p(\mathbf{x}_{i-1:i} | \mathbf{z}_{0:k}) \\ &= \int p(\mathbf{x}_{0:i} | \mathbf{z}_{0:k}) d\mathbf{x}_{0:i-2} = Z_k^{-1} \int p(\mathbf{x}_{0:i}, \mathbf{z}_{0:i}) d\mathbf{x}_{0:i-2} d\mathbf{x}_{i+1:k} \\ &= Z_k^{-1} p(\mathbf{z}_{i-1} | \mathbf{x}_{i-1}) p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{z}_i | \mathbf{x}_i) \\ &\quad \times \int p(\mathbf{x}_{0:i-2}, \mathbf{z}_{0:i-2}) p(\mathbf{x}_{i-1} | \mathbf{x}_{i-2}) d\mathbf{x}_{0:i-2} \int \prod_{j=i+1}^k p(\mathbf{x}_j | \mathbf{x}_{j-1}) p(\mathbf{z}_j | \mathbf{x}_j) d\mathbf{x}_{i+1:k} \\ &\leq p(\mathbf{x}_{i-1} | \mathbf{z}_{0:i-1}) (C_1 C_2)^{-1} C_3 p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{z}_i | \mathbf{x}_i). \end{aligned}$$

The last inequality holds because, according to the definition of $p(\mathbf{x}_{i-1} | \mathbf{z}_{0:i-1})$,

$$p(\mathbf{x}_{i-1} | \mathbf{z}_{0:i-1}) = Z_{i-1}^{-1} \int p(\mathbf{x}_{0:i-2}, \mathbf{z}_{0:i-2}) p(\mathbf{x}_{i-1} | \mathbf{x}_{i-2}) d\mathbf{x}_{0:i-2} p(\mathbf{z}_{i-1} | \mathbf{x}_{i-1}),$$

and (31) holds from condition (28). It therefore follows that

$$\frac{p(\mathbf{x}_{i-1:i}|\mathbf{z}_{0:k})}{p(\mathbf{x}_{i-1}|\mathbf{z}_{0:i-1})q(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{z}_{0:i})} \leq (C_1C_2)^{-1}C_3C_4, \quad (41)$$

where the last inequality holds because of condition (39).

Hence, from (40), (41), and the fact that if $\psi(\mathbf{x}, \mathbf{y}) \leq C_u$ and g is a density function

$$\left| \int \psi(\mathbf{x}, \mathbf{y})g(\mathbf{x}|\mathbf{z})d\mathbf{x} \right| \leq C_u \int g(\mathbf{x}|\mathbf{z})d\mathbf{x} = C_u,$$

each component of variance V_k in (25) is upper bounded by $4(C_1C_2)^{-1}C_3C_4\|\phi\|^2(1 - C^{-2})^{2k-2i}$. This implies that

$$\begin{aligned} V_k &\leq 4(C_1C_2)^{-1}C_3C_4\|\phi\|^2 \sum_{i=0}^k (1 - C^{-2})^{2k-2i} \\ &\leq 4(C_1C_2)^{-1}C_3C_4\|\phi\|^2 \frac{1}{1 - (1 - C^{-2})^2}. \end{aligned}$$

The proof is then complete. \square

To summarize, we have the following theorem.

Theorem 1. *Suppose that $\|\phi\| = \max \phi \leq \infty$, $\boldsymbol{\mu}_i$ in (1) and $\boldsymbol{\nu}_i$ in (2) are truncated normal distributed and the proposal density is chosen as in (24). Then*

$$N^{1/2}(I^{\text{smc}}(\phi, k, p) - I(\phi, k, p))$$

converges to $\mathcal{N}(0, V_k)$ in distribution, where V_k increases linearly with k if $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_{0:k})$ and is independent of k if $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_k)$.

Proof. Because $\boldsymbol{\mu}, \boldsymbol{\nu}$ are truncated normal distributed and the proposal density is chosen as in (24), conditions (28) and (29) are satisfied for $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_{0:k})$. According to Lemma 1, V_k increases linearly with k . Note that conditions (28),(37), and (39) hold for truncated normal density functions. That is, the conditions in Lemma 2 apply. Hence V_k is independent of k . The proof is complete. \square

3.5. Numerical Convergence Rate

Assume that the operators M and H in the definition of the state space model (1)–(2) are not easy to compute. That situation happens, for example, if one needs to integrate a differential equation between the observations.

We then have to compute with an approximate state space model such as

$$\mathbf{x}_i = \tilde{M}(\mathbf{x}_{i-1}) + \boldsymbol{\mu}_i, \quad (42)$$

$$\mathbf{z}_i = \tilde{H}(\mathbf{x}_i) + \boldsymbol{\nu}_i. \quad (43)$$

Then we have $\tilde{p}(\mathbf{x}_i|\mathbf{x}_{i-1}) = p_{\boldsymbol{\mu}_i}(\mathbf{x}_i - \tilde{M}(\mathbf{x}_{i-1}))$, and the conditional density of \mathbf{z}_i given \mathbf{x}_i is $\tilde{p}(\mathbf{z}_i|\mathbf{x}_i) = p_{\boldsymbol{\nu}_i}(\mathbf{z}_i - \tilde{H}_i(\mathbf{x}_i))$. And the posterior is denoted as $\tilde{p}(\mathbf{x}_{0:k}|\mathbf{z}_{0:k})$. To make sure that SMC methods still produce reasonable results for good approximations of M and H , we require that $\log(p_{\boldsymbol{\mu}_i}), \log(p_{\boldsymbol{\nu}_i})$ are Lipschitz continuous and hence have the following theorem.

Theorem 2. Assume that the conditions of Theorem 1 are satisfied and $\log(p_{\mu_i}), \log(p_{\nu_i})$ are Lipschitz continuous with Lipschitz coefficient L_μ and L_ν respectively. If furthermore $|\tilde{M}(\mathbf{x}) - M(\mathbf{x})| < \epsilon/(2L_\mu)$, $|\tilde{H}(\mathbf{x}) - H(\mathbf{x})| < \epsilon/(2L_\nu)$ and ϵ is small enough, then for sequential Monte Carlo methods using multinomial resampling, we have that

$$I^{\text{smc}}(\phi, k, \tilde{p}) - I(\phi, k, p)$$

converges to $\mathcal{N}(E_k, N^{-1}V_k)$ in distribution, where V_k is linear with k if $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_{0:k})$, V_k is independent of k if $\phi : \mathbf{x}_{0:k} \rightarrow \phi(\mathbf{x}_k)$, and E_k is upper bounded by $2\|\phi\| (\exp((k+1)\epsilon) - 1)$.

Proof. We can rewrite $I^{\text{smc}}(\phi, k, \tilde{p}) - I(\phi, k, p)$ as

$$I^{\text{smc}}(\phi, k, \tilde{p}) - I(\phi, k, p) = (I^{\text{smc}}(\phi, k, \tilde{p}) - I(\phi, k, \tilde{p})) + (I(\phi, k, \tilde{p}) - I(\phi, k, p)).$$

The first component is already discussed in Theorem 1. Only the second component is needed for the discussion in this proof. The second parts describe the bias introduced by using approximations of H and M .

Let

$$\begin{aligned} T_i := & \log\left(p_{\mu_i}(\mathbf{x}_i - \tilde{M}(\mathbf{x}_{i-1}))\right) - \log\left(p_{\mu_i}(\mathbf{x}_i - M(\mathbf{x}_{i-1}))\right) + \\ & \log\left(p_{\nu_i}(\mathbf{z}_i - \tilde{H}(\mathbf{x}_i))\right) - \log\left(p_{\nu_i}(\mathbf{z}_i - H(\mathbf{x}_i))\right). \end{aligned} \quad (44)$$

Because $\log(p_{\mu_i}), \log(p_{\nu_i})$ are Lipschitz continuous and $|\tilde{M}(\mathbf{x}) - M(\mathbf{x})| < \epsilon/(2L_\mu)$, $|\tilde{H}(\mathbf{x}) - H(\mathbf{x})| < \epsilon/(2L_\nu)$, we have that that

$$|T_i| \leq L_\mu \frac{\epsilon}{2L_\mu} + L_\nu \frac{\epsilon}{2L_\nu} = \epsilon. \quad (45)$$

Using this inequality, we can prove that

$$\begin{aligned} |\tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) - p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k})| & \leq p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) \left(\exp\left(\sum_{i=0}^k |T_i|\right) - 1 \right) \\ & \leq (\exp((k+1)\epsilon) - 1) p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}). \end{aligned} \quad (46)$$

In the same fashion, we can get

$$|\tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) - p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k})| = (\exp((k+1)\epsilon) - 1) \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}). \quad (47)$$

It is therefore easy to prove that

$$\begin{aligned} S_1 & := \left| \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) \int (p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) - \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k})) d\mathbf{x}_{0:k} \right| \\ & \leq (\exp((k+1)\epsilon) - 1) \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) \int p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{0:k}, \end{aligned} \quad (48)$$

and

$$\begin{aligned} S_2 & := \left| (\tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) - p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k})) \int \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{0:k} \right| \\ & \leq (\exp((k+1)\epsilon) - 1) p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) \int \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{0:k}. \end{aligned} \quad (49)$$

Hence it is not difficult to prove that

$$\begin{aligned}
& |I(\phi, k, \tilde{p}) - I(\phi, k, p)| \\
&= \left| \int \phi(\mathbf{x}_{0:k}) (\tilde{p}(\mathbf{x}_{0:k}|\mathbf{z}_{0:k}) - p(\mathbf{x}_{0:k}|\mathbf{z}_{0:k})) d\mathbf{x}_{0:k} \right| \\
&\leq \|\phi\| \int |\tilde{p}(\mathbf{x}_{0:k}|\mathbf{z}_{0:k}) - p(\mathbf{x}_{0:k}|\mathbf{z}_{0:k})| d\mathbf{x}_{0:k} \\
&\leq \|\phi\| \frac{\int (S_1 + S_2) d\mathbf{x}_{0:k}}{\int \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{0:k} \int p(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{0:k}}. \tag{50}
\end{aligned}$$

To obtain the last inequality, we rewrite $\tilde{p}(\mathbf{x}_{0:k}|\mathbf{z}_{0:k})d\mathbf{x}_{0:k}$ and $p(\mathbf{x}_{0:k}|\mathbf{z}_{0:k})d\mathbf{x}_{0:k}$, according to the fact that

$$g(\mathbf{x}_{0:k}|\mathbf{z}_{0:k}) = \frac{g(\mathbf{x}_{0:k}, \mathbf{z}_{0:k})}{\int g(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{0:k}}$$

for any conditional density g , and then combine the two fractions into one. After that, in the numerator we add $(\int \tilde{p}(\mathbf{x}_{0:k}, \mathbf{z}_{0:k}) d\mathbf{x}_{0:k})^2$, and then subtract it, and apply triangle inequality for absolute values. From (48), (49), and (50), it follows that

$$|I(\phi, k, \tilde{p}) - I(\phi, k, p)| \leq 2\|\phi\| (\exp((k+1)\epsilon) - 1).$$

The proof is complete. \square

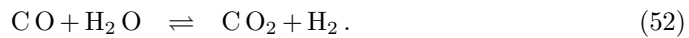
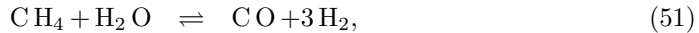
Note that if $(k+1)\epsilon$ is small, $\exp((k+1)\epsilon) - 1$ is close to $(k+1)\epsilon$. Hence, the bias is accumulated linearly with k when k is in a moderately large region if accurate approximations of M and H are used.

Theorem 2 shows that sequential Monte Carlo methods can give us reasonably good estimations if the log densities of μ_i and ν_i are Lipschitz continuous and approximations of M and H are good enough. This convergence result is similar to Theorem 1, except that the errors in the approximations introduce a bias and the bias will increase with the number of steps. Hence, to get estimations of fixed accuracy, we need to increase the accuracy of approximation when the number of steps increases.

In the test example in next section, $M(\mathbf{x})$ and $H(\mathbf{x})$ are functions of the solution of an ordinary differential equation (ODE) system with \mathbf{x} as the initial value. During the simulation, we can get only an approximation of $M(\mathbf{x})$ or $H(\mathbf{x})$ by numerical schemes. Theorem 2 shows that SMC methods would work well with such problem if a good numerical scheme for the initial-valued ODE is provided during the simulation.

4. Numerical Demonstration of the Approach

We now present details for numerical simulations for the steam methane reforming (SMR) process. The chemical reactions involved are



We model this chemical process according to [29] and build the ODE system to describe the reaction according to [30]. For details on the setup, see Appendix A.

4.1. Sequential Importance Sampling Toolkit for ODE Systems

We developed a library called Sequential Important Sampling Toolkit for ODE Systems to produce samples from the HMM posterior using a particle method. SISTOS provides the following

1. Importance sampling, sequential importance sampling and sequential Monte Carlo sampling. Proposal densities used are the normal and the truncated normal.
2. Systematic, residual and multinomial resampling methods [22, 21] for the sequential Monte Carlo method.
3. Parallel programming support by means of the PETSc library’s [31] syntax and data structure. Note that resampling requires a synchronization step at each time step. Therefore parallelism is not difficult to implement, but it is not trivial either.
4. Automatic differentiation support (for use in the linearization and maximum likelihood procedures) through the ADIC [32] and ADOLC [33] libraries.
5. Optimization support (for use in the maximum likelihood procedure) through TAO [34].

4.2. Performance result

In the numerical experiments, we compute the average amount of leakage of CO, CO₂ and H₂. For each cycle, t_f is set to be 1. For the random variable μ_i , we take $Q_i = 0.01I$ and for ν_i , $R_i = 0.01$. The upper bound of the random number $\mu_{\text{H}_2\text{O},i}$ is 6 and of $\mu_{\text{C H}_4}$ is 2, and their lower bound is 0. For ν_i , its upper bound is 1 and its lower bound is -1 . All tests are run on the high-performance cluster Fusion at Argonne National Laboratory. We plot numerical results in loglog scale. We can see from Figure 1(b) and 1(d) that the run time is increasing approximately linearly with the number of sample points, which indicates that the targeted observable is predictable, as stated in Theorem 1 and Theorem 2.

We first test a five-step SMR example in 20 independent runs for a comparison between importance sampling (IS) and SMC. The reason we only choose a five-step SMR example here is that for a 10-step example would fail for the IS method. The proposal density of IS is chosen here by the Laplace’ method [35]. We use the maximum likelihood method to choose proposal densities and systematic resampling technique for SMC. From Figure 1(a), we can see that SMC performs much better than IS.

We test the 100-step and SMR example in 100 independent runs using eight processors to compare the maximum likelihood method and the linearization method.

In Figure 1(c), we compare the numerical results of the maximum likelihood and the linearization method. We can see that both methods show a convergence rate $\mathcal{O}(N^{-1/2})$ approximately, as predicted by the theory.

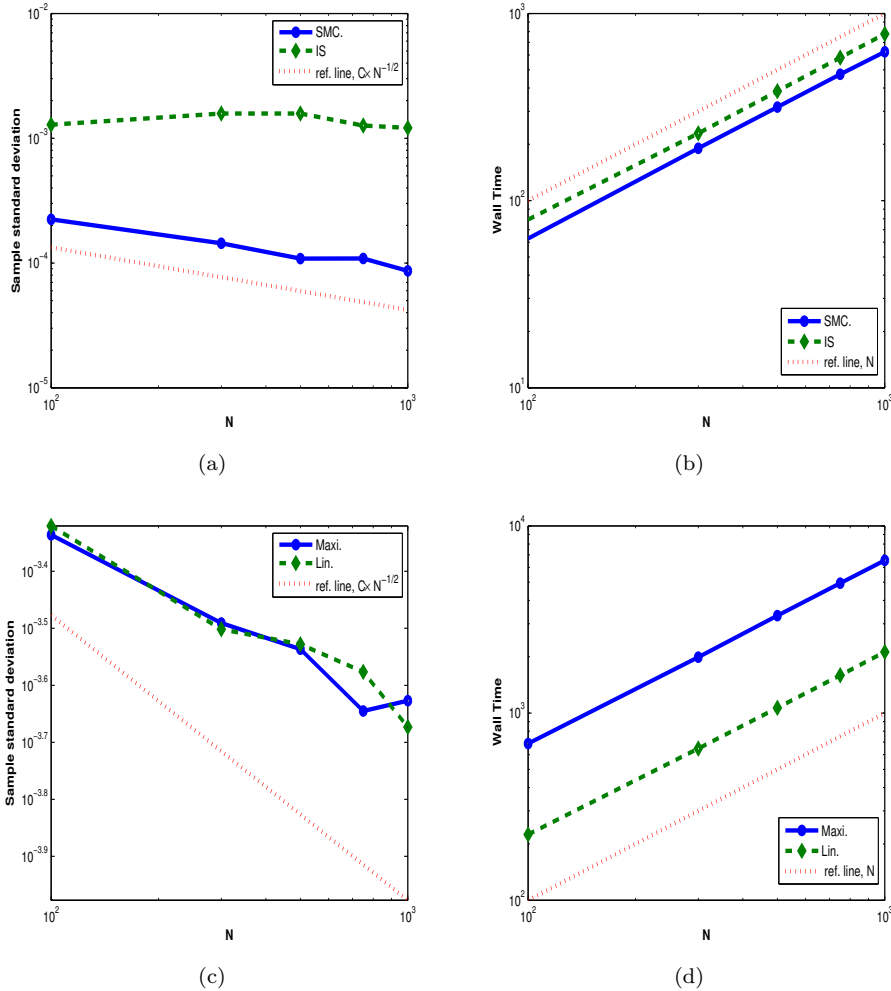


Figure 1: Numerical Results: (a) comparison of sample standard deviation between IS and SMC when $k = 5$. SMC uses systematic resampling method; (b) comparison of wall time between IS and SMC when $k = 5$. SMC uses systematic resampling method; (c) comparison of sample standard deviation between maximum likelihood method and linearization method when $k = 100$. Use systematic resampling; (d) comparison of wall time deviation between maximum likelihood method and linearization method when $k = 100$. Use systematic resampling.

4.3. Comparison between the maximum likelihood or linearization SMC approach and the bootstrap SMC approach

In our theoretical developments in §3 we were concerned with finding conditions for the maximum likelihood or linearization filters to avoid the failure of the exponential type as was encountered by the bootstrap particle filter [27]. The latter is an SMC approach that uses as a proposal density $p(x_i|x_{i-1})$ as opposed to the one obtained by maximum likelihood or linearization as we have derived here. It can be proven that,

under assumptions essentially similar to the ones of our results in §3, the bootstrap particle filter also avoids the exponential failure in this case.

However, the bootstrap particle filter does have a shortcoming compared to the maximum likelihood or linearization SMC, that is, its proposal density does not accommodate observation noise, as it is tuned exclusively to the dynamical part.

Let's take 1-d state space model for example.

$$x_i = 0.5x_{i-1} + 25x_{i-1}/(1 + x_{i-1}^2) + 8 \cos(1.2i) + \sigma_\mu \mu_i; \quad (53)$$

$$z_i = x_i + \sigma_\nu \nu_i. \quad (54)$$

Here $\mu_i \sim \mathcal{N}(0, 1)$ and $\nu_i \sim \mathcal{N}(0, 1)$. It can be immediately shown that the linearization method and maximum likelihood method introduced in this paper result in the same proposal distribution as $\mathcal{N}(\bar{\mu}_i, \sigma_i^2)$, with $\sigma_i^2 = (\sigma_\mu^{-2} + \sigma_\nu^{-2})^{-1}$ and $\bar{\mu}_i = \sigma_i^2(z_i/\sigma_\nu^2 + (0.5x_{i-1} + 25x_{i-1}/(1 + x_{i-1}^2) + 8 \cos(1.2i))/\sigma_\mu^2)$.

We test a 10-step 1-d example in 20 independent runs for a comparison of the standard deviations between the bootstrap particle filter and maximum likelihood (linearization) particle filter. In Fig.3, different colors represent different σ_μ (state noise level) and different markers represent different particle filters. From Fig.3, we can see that the performance of the maximum likelihood particle filter is relatively stable with increasing σ_μ , whereas the one of the bootstrap particle filter significantly degrades with increasing σ_μ , as was also suggested by the argument above. We get some insights of the cause for this phenomenon from Figure 2. With the σ_μ increasing, the proposal density of the bootstrap becomes flatter, as the dynamical part becomes less informative. On the other hand, the shape of the proposal density based on the maximum likelihood or linearization approach does not change much because it also incorporates the effect of observations whose noise level is kept constant, see Fig. 2. When applications have a weak dynamic model but moderately accurate observations, the bootstrap density is not sufficiently informative, which in turn results in the high variance of the estimates as indicated in Figure 3.

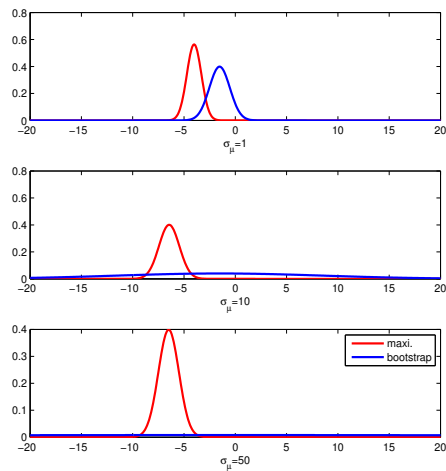


Figure 2: The importance density plots of the bootstrap particle filter and maximum likelihood method in 1-d example

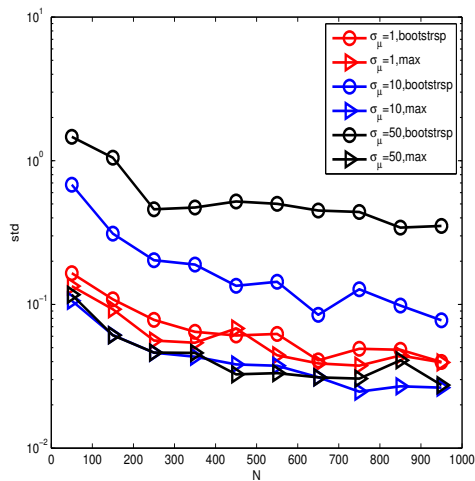


Figure 3: Comparison between the bootstrap particle filter and maximum likelihood method in 1-d example

5. Conclusion

In this article, we investigate the behavior of sequential Monte Carlo methods when used for the estimation of state space uncertainty in nonlinear dynamical systems. The uncertainty model used is the hidden Markov model associated with differential equations. Our main objective is to investigate whether such methods are tractable for assessing uncertainty in dynamical systems of the type encountered in chemical engineering applications. We investigate the case where the proposal densities are produced by either the linearization method or conditional maximum likelihood method. We prove that for two state functionals widely encountered in applications, the scaled variance of the resulting statistical estimator behaves favorably. That is, the estimates produced by sequential Monte Carlo methods converge in distribution, and the variance is confined to a linear increase with an increase on the number of steps for accumulated estimates and is bounded by a constant for terminal estimates. Thus, the method is tractable in these circumstances.

We are also interested in the practical case where the computation of the dynamical and observational maps cannot be carried out exactly. We show that errors arising from computing M and H introduce a bias that depends on the accuracy with which these mappings are computed. The bias accumulates with an increase in the number of steps. Hence, we need to improve the accuracy of the approximation of H and M if the number of steps increases.

For numerical demonstrations, we use the steam methane reforming process as an example to demonstrate the SMC behavior on dynamical systems originating from chemical systems. The numerical results are consistent with the theoretical results. The approach has been implemented in the parallel portable library SISTOS.

Acknowledgments

We are grateful to the anonymous referee for comments that have improved the paper, particularly for the suggestion of describing differences with the bootstrap filter. We are grateful to Candido Pereira and John Krebs from the Chemical Sciences and Engineering Division at Argonne National Laboratory for comments and suggestions concerning applications in chemical engineering. This work was supported by the U.S. Department of Energy under contract DE-AC02-06CH11357.

References

- [1] W. Bebbington, Reprocessing of nuclear fuels, *Sci. Am.* 235 (1976) 30–41.
- [2] G. Uchiyama, S. Fujine, S. Hotoku, M. Maeda, New separation process for neptunium, plutonium, and uranium using butyraldehydes as reductants in reprocessing, *Nuclear Technology* 102 (1993) 341–352.
- [3] K. Crowley, Nuclear waste disposal: the technical challenges, *Physics Today* 50 (1997) 32–40.
- [4] J. Birkett, M. Carrott, O. Fox, C. Jones, C. Maher, C. Roube, R. Taylor, D. Woodhead, Recent developments in the Purex process for nuclear fuel reprocessing: Complexant based stripping for uranium/plutonium separation, *CHIMIA International Journal for Chemistry* 59 (2005) 898–904.
- [5] L. Rabiner, B. Juang, An introduction to hidden Markov models, *IEEE ASSp Magazine* 3 (1986) 4–16.
- [6] G. Welch, G. Bishop, An introduction to the Kalman filter, University of North Carolina at Chapel Hill, Chapel Hill, NC (1995).

- [7] P. Houtekamer, H. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review* 126 (1998) 796–811.
- [8] P. Houtekamer, H. Mitchell, A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review* 129 (2001) 123–137.
- [9] E. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in: *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000 (AS-SPCC)*, IEEE, pp. 153–158.
- [10] E. Kalnay, *Atmospheric Modeling, Data Assimilation, and Predictability*, Cambridge Univ. Press, 2003.
- [11] R. Mamon, R. Elliott, *Hidden Markov Models in Finance*, Springer Verlag, 2007.
- [12] U. Marti, H. Bunke, Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition systems, *World Scientific Series in Machine Perception and Artificial Intelligence Series* (2001) 65–90.
- [13] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* 77 (1989) 257–286.
- [14] A. Schliep, A. Schonhuth, C. Steinhoff, Using hidden Markov models to analyze gene expression time course data, *Bioinformatics* 19 (2003) i255–i263.
- [15] J. Hammersley, K. Morton, Poor man’s Monte Carlo, *Journal of the Royal Statistical Society. Series B (Methodological)* 16 (1954) 23–38.
- [16] M. Rosenbluth, A. Rosenbluth, Monte Carlo calculation of the average extension of molecular chains, *Journal of Chemical Physics* 23 (1955) 356–359.
- [17] H. Akashi, H. Kumamoto, Construction of discrete-time nonlinear filter by Monte Carlo methods with variance-reducing techniques, *Systems and Control* 19 (1975) 211–221.
- [18] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research* 99 (1994) 10143.
- [19] K. Kremer, K. Binder, Monte Carlo simulation of lattice models for macromolecules, *Computer Physics Reports* 7 (1988) 259–310.
- [20] M. Evans, T. Swartz, Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems, *Statistical Science* 10 (1995) 254–272.
- [21] J. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2001.
- [22] A. Doucet, A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later, *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press. To appear (2009).
- [23] A. Hindmarsh, R. Serban, User documentation for CVODES, an ODE solver with sensitivity analysis capabilities, Lawrence Livermore National Laboratory 189, report no. UCRL-MA-148813 (2002).
- [24] N. Chopin, Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference, *Annals of Statistics* 32 (2004) 2385–2411.
- [25] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer Verlag, 2004.
- [26] H. Kunsch, Recursive Monte Carlo filters: Algorithms and theoretical analysis, *Annals of Statistics* 33 (2005) 1983–2021.
- [27] P. Bickel, B. Li, T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions, in: *IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3, pp. 318–329.
- [28] R. Dobrushin, Central limit theorem for non-stationary markov chains i, ii, *Theory Probab. Appl.* 1 (1956) 65–80, 329–383.
- [29] X. Zeng, M. Anitescu, C. Pereira, M. Regalbuto, A Framework for chemical plant safety assessment under uncertainty, *Studies in Informatics and Control* 18 (2009) 7–20.
- [30] C. Singh, D. Saraf, Simulation of side-fired steam-hydrocarbon reformers, *Ind. Eng. Chem. Process Des. Dev.* 18 (1979) 1–7.
- [31] S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, H. Zhang, *PETSc Users Manual*, Technical Report ANL-95/11 - Revision 3.1, Argonne National Laboratory, 2010.
- [32] C. Bischof, L. Roh, A. Mauer-Oats, ADIC: an extensible automatic differentiation tool for ANSI-C, *Urbana* 51 (1997) 61802.
- [33] A. Griewank, D. Juedes, J. Srinivasan, et al., ADOL-C, a package for the automatic differentiation of algorithms written in C/C++, *ACM Trans. Math. Software* 22 (1996).
- [34] S. Benson, L. C. McInnes, J. Moré, T. Munson, J. Sarich, *TAO User Manual (Revision 1.9)*, Technical Report ANL/MCS-TM-242, Mathematics and Computer Science Division, Argonne National

- Laboratory, 2007. [Http://www.mcs.anl.gov/tao](http://www.mcs.anl.gov/tao).
- [35] B. Carlin, T. Louis, Bayes and empirical Bayes methods for data analysis, *Statistics and Computing* 7 (1997) 153–154.

The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory (“Argonne”) under Contract No. DE-AC02-06CH11357 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

Appendix A. Steam Methane Reforming

In this appendix, we use the steam methane reforming (SMR) process as an example to demonstrate the setup of the hidden Markov model. SMR is used in industries to separate hydrogen atoms from carbon atoms in methane.

We model the chemical process in the following way [29]: At each end of the cycle, the products are extracted, and the unconsumed reactants are combined with the new reactant feed.

We define a vector $\mathbf{Y}_i \in \mathcal{Y} \subset \mathcal{R}^{n_y}$ that has two parts,

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_i^{in} \\ \mathbf{Y}_i^{out} \end{bmatrix}.$$

Here \mathbf{Y}_i^{in} denotes the recycled components and \mathbf{Y}_i^{out} denotes the extracted components. We describe the chemical process as an initial-valued ODE

$$\dot{\mathbf{Y}}_i = F(\mathbf{Y}_i), \quad (\text{A.1})$$

$$\mathbf{Y}_i^{in}(0) = \mathbf{X}_i, \quad (\text{A.2})$$

$$\mathbf{Y}_i^{out}(0) = 0. \quad (\text{A.3})$$

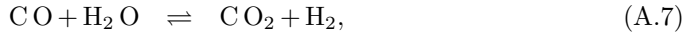
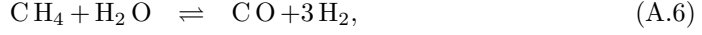
Here \mathbf{X}_i is a vector of state variables, which denotes the initial amounts of reactants at each cycle. We compute the solution after t_f , $\mathbf{Y}_i(t_f)$ from ODE system, and we define

$$M(\mathbf{X}_i) := \mathbf{Y}_i^{in}(t_f), \quad (\text{A.4})$$

$$H(\mathbf{X}_i) := h_i(\mathbf{Y}_i^{out}(t_f)). \quad (\text{A.5})$$

This completes the definition of the HMM components (1)–(2).

The SMR reaction is



We denote

$$\mathbf{Y}^{in} = [y_{\text{H}_2\text{O}}, y_{\text{C H}_4}]^T, \mathbf{Y}^{out} = [y_{\text{C O}}, y_{\text{C O}_2}, y_{\text{H}_2}]^T, \text{ and } \mathbf{X}_i = [x_{\text{H}_2\text{O},i}, x_{\text{C H}_4,i}]^T.$$

Let \mathcal{I} denote the character set $\{\text{H}_2\text{O}, \text{CH}_4, \text{CO}, \text{CO}_2, \text{H}_2\}$. The reaction equations [30] chosen here for SMR are

$$r_1 = a_1 \exp(-E_{a_1}/RT) \sqrt{RT} \left(\frac{y_{\text{C H}_4}}{\sqrt{\sum_{j \in \mathcal{I}} y_j}} - \frac{y_{\text{C O}} y_{\text{H}_2}^3 (RT)^2}{k_{eq1} y_{\text{H}_2\text{O}} \sqrt{\sum_{j \in \mathcal{I}} y_j}} \right), \quad (\text{A.8})$$

$$r_2 = a_2 \exp(-E_{a_2}/RT) \sqrt{RT} \left(\frac{y_{\text{C O}}}{\sqrt{\sum_{j \in \mathcal{I}} y_j}} - \frac{y_{\text{C O}_2} y_{\text{H}_2}}{k_{eq2} y_{\text{H}_2\text{O}} \sqrt{\sum_{j \in \mathcal{I}} y_j}} \right). \quad (\text{A.9})$$

Hence the right-hand side of the ODE is

$$F(\mathbf{Y}) = \begin{pmatrix} -r_1 - r_2 \\ -r_1 \\ r_1 - r_2 \\ r_2 \\ 3r_1 + r_2 \end{pmatrix}. \quad (\text{A.10})$$

Let $y_{j,i}, j \in \mathcal{I}, i = 0, \dots, k$ be the solution of the ODE system defined by (A.1)–(A.3) and (A.8)–(A.10) at $t_{i+1} = t_i + t_f$. Hence

$$M(x_{\text{H}_2\text{O},i}, x_{\text{CH}_4,i}) := \begin{cases} y_{\text{H}_2\text{O},i}, \\ y_{\text{CH}_4,i}, \end{cases} \quad (\text{A.11})$$

and

$$H(x_{\text{H}_2\text{O},i}, x_{\text{CH}_4,i}) := y_{\text{H}_2,i} + y_{\text{CO},i} + y_{\text{CO}_2,i}. \quad (\text{A.12})$$

The hidden Markov model thus is

$$x_{\text{H}_2\text{O},i+1} = y_{\text{H}_2\text{O},i} + \mu_{\text{H}_2\text{O},i+1} \quad (\text{A.13})$$

$$x_{\text{CH}_4,i+1} = y_{\text{CH}_4,i} + \mu_{\text{CH}_4,i+1}, \quad (\text{A.14})$$

$$z_{i+1} = y_{\text{H}_2,i+1} + y_{\text{CO},i+1} + y_{\text{CO}_2,i+1} + \nu_i. \quad (\text{A.15})$$

Here $\mu_{i,\text{H}_2\text{O}}$ and μ_{i,CH_4} are random variables representing the amounts of refills, and ν_i denotes the uncertainty of the observation.