Particle Filters for the Estimation of a State Space Model

Tao Chen*, Julian Morris and Elaine Martin
Centre for Process Analytics and Control Technology
School of Chemical Engineering and Advanced Materials
University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK

Abstract
In process engineering, on-line state estimation is a key element in state space modelling. However when state/measurement functions are highly non-linear and the posterior probability of the state is non-Gaussian, conventional filters, such as the extended Kalman filter, do not provide satisfactory results. This paper proposes an alternative approach whereby particle filters based on the sequential Monte Carlo method are used for the estimation task. Particle filters are described prior to discussing a number of implementation issues. Furthermore, a Markov chain Monte Carlo method is proposed to enhance particle filters where the estimates of the initial conditions are poor. The effectiveness of particle filters is demonstrated through the state estimation of an exothermic irreversible parallel reaction in a continuous stirred tank reactor.

Keywords: Markov chain Monte Carlo, particle filter, sequential Monte Carlo, state space model.

1. Introduction
In process engineering, the assured on-line estimation of a dynamic state space model is a key element of model based approaches, such as optimization, performance monitoring and process control. State estimation can be viewed as an optimal filtering problem under a Bayesian framework. If the state equations are linear and the posterior density is Gaussian, the Kalman filter (KF) (Jazwinski, 1970) provides an optimal solution. However, where these assumptions do not hold, there exists no analytical solution and therefore approximations have to be made. One example is the extended Kalman filter (EKF) which assumes a Gaussian posterior density and adopts a first-order Taylor expansion to provide local approximation within the current state. In practice where the state equations are highly non-linear and the posterior density is non-Gaussian, the EKF may give a large estimation error. An alternative algorithm is the approximate grid-based filter which was introduced into process engineering by Terwiesch (1995). The computational cost of the grid-based filter increases exponentially with the state dimension, thus limiting its widespread application.

In this paper, particle filters are introduced for dynamic state space estimation. The basic idea behind particle filters is to approximate the posterior probability of the states

* To whom correspondence should be addressed: tao.chen@newcastle.ac.uk
using a large number of samples (particles) with associated weights. These particles and weights are then updated sequentially along with the state evolution when new observations become available. A particle filter, namely the Auxiliary Sequential Importance Re-sampling (ASIR) filter (Pitt and Shephard, 1999) is described. When efficiently implemented, the ASIR filter incurs computational costs proportional to the number of particles, thus satisfying the requirements for on-line estimation when conventional computational facilities are available.

However, when the initial conditions deviate far from the true value, the probability of the particles approaching the true states can be very small in the early stages. Consequently particle filters can converge slowly. A hybrid solution is to use the EKF to move the particles to the posterior density at each time step (De Freitas et al., 2000). An alternative solution based on the Markov Chain Monte Carlo (MCMC) method is proposed. A Gibbs sampler with an accept/reject step is adopted to generate samples that approximate the posterior probability of the states. As the computational cost of MCMC increases significantly with time, MCMC is only performed for the first few time steps before transferring over to the conventional particle filter.

The effectiveness of particle filters is demonstrated through the state space estimation of an exothermic irreversible parallel reaction taking place within a continuous stirred tank reactor (CSTR). For good estimates of the initial conditions, the particle filter is shown to provide significantly lower estimation errors than the EKF. In the case of very poor initial guesses, the proposed MCMC method provides faster convergence.

### 2. An Overview of Particle Filters

A state space model is defined by the following state and measurement functions:

\[
\begin{align*}
    x_k & = f_k(x_{k-1}, v_{k-1}) \\
    z_k & = h_k(x_k, n_k)
\end{align*}
\]  

(1)

where \( k \) is the time index and \( f_k \) is a non-linear function describing the evolution of the state with independent and identically distributed process noise, \( v_{k-1} \). \( h_k \) is a non-linear function mapping the state space to the measurements with independent and identically distributed noise \( n_k \). Let \( z_{1:k} = (z_i, i = 1, \ldots, k) \) be the measurement sequence and assuming the prior distribution \( p(x_0) \) is known, the posterior probability can be obtained sequentially by prediction and updated as follows:

\[
\begin{align*}
    p(x_k \mid z_{1:k-1}) & = \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid z_{1:k-1}) dx_{k-1} \\
    p(x_k \mid z_{1:k}) & = \frac{p(z_k \mid x_k) p(x_k \mid x_{k-1})}{p(z_k \mid z_{1:k-1})}
\end{align*}
\]  

(2)

The above equations are the optimal solution from a Bayesian perspective to the non-linear state estimation problem. One limitation is that the evolution of the posterior density cannot in general be determined analytically. Thus some approximation must be
Particle filters approximate the posterior probability by a set of support points (particles) \( x_k^i, i = 1,...,N \), with associated weights \( w_k^i \):

\[
p(x_k \mid z_{1:k}) = \sum_{i=1}^{N} w_k^i \delta(x_k - x_k^i)
\]

where \( \delta(.) \) is an indicator function. In this paper the filtered state is taken as the mean of the posterior density. The weights are decided using importance sampling (Robert and Casella, 1999). An importance density \( q(x_k \mid z_{1:k}) \) is identified from which samples are drawn. The weights are then defined as being proportional to the ratio of \( p(x_k \mid z_{1:k}) \) to \( q(x_k \mid z_{1:k}) \). It has been shown (Arulampalam et al., 2002) that if the importance density is selected appropriately and is only dependent on the current observation, \( z_k \), and the past state, \( x_{k-1} \), the weights can be updated as follows:

\[
w_k^i \propto w_{k-1}^i \frac{p(z_k \mid x_k^i)p(x_k^i \mid x_{k-1}^j)}{q(x_k^i \mid x_{k-1}^j, z_k)}
\]

Two implementation issues are considered. The first is that of degeneracy, where after some time steps, only one particle has significant weight. Thus considerable computational effort will have been spent updating particles whose contribution to the approximation of \( p(x_k \mid z_{1:k}) \) is negligible. Re-sampling can be used to eliminate those particles with small weights thereby focussing on particles with large weights. Resampling generates a new particle set by sampling with replacement from the original set \( \{ x_k^i, i = 1,...,N \} \) with \( \text{Pr}(x_k^i = x_k^j) = w_k^j \). Here \( j \) is the particle index after resampling. The parent relationship is denoted, \( \text{parent}(j) = i \). The weights are reset to \( 1/N \) as the samples are independent and identically distributed and are drawn from a discrete density function.

The second issue is how to choose the importance density. A convenient choice is to use the prior, \( p(x_k \mid x_{k-1}^i) \). A particle filter with this importance density and re-sampling step is called a sequential importance re-sampling (SIR) filter (Gordon et al., 1993). However, as the importance density is independent of the current measurement, the state space is explored without knowledge of the observations, which makes the SIR filter sensitive to outliers. Recently the Auxiliary SIR (ASIR) filter (Pitt and Shephard, 1999) was proposed to obtain a more reliable importance density. It is defined as:

\[
q(x_k^i \mid z_k) \propto p(z_k \mid \mu_k^i)p(x_k \mid x_{k-1}^i)w_{k-1}^i
\]

where \( \mu_k^i \) can be the mean of \( x_k \) conditional on \( x_{k-1}^i \) or an associated sample. In addition Bayes's rule shows that:
Considering the re-sampling step presented earlier, the particle $x^j_k$ is assigned a weight proportional to the ratio of the right-hand side of equation (6) to equation (5) as:

$$w^j_k \propto \frac{p(z_k \mid x^j_k)}{p(z_k \mid \mu^\text{parent}(j))}$$

The ASIR filter generates particles from the sample at time step $k-1$ conditional on the current observation, which can be closer to the true state compared with those obtained using the SIR filter. If the noise of state evolution is low, even with relatively high measurement noise, ASIR is less likely to be sensitive to outliers as $p(x_k \mid x^j_{k-1})$ is well characterized by $\mu^j_k$.

3. State Space Estimation with Poor Initial Conditions

Good initial conditions, specified through the prior distribution, can guarantee rapid convergence of the particle filtering approaches. However, in practice, when the initial conditions deviate far from the true values, the probability of the particles approaching the true states can be very low in the early stages. A hybrid solution has been proposed (De Freitas et al., 2000) to speed up the convergence of particle filters. This approach first generates particles as for conventional particle filters. It then tries to move particles closer to the posterior distribution through an EKF step. Although this strategy was successfully applied for the training of neural networks, it faces the same problems as the EKF, i.e. it can work poorly when the assumptions for the EKF do not hold.

In this paper an alternative solution based on the Markov Chain Monte Carlo (MCMC) method is proposed. For an introduction to MCMC methods, see, for example, Robert and Casella (1999). In each time step, $k$, a state sequence $\{x_i, I = 0:k\}$ is maintained. The objective is to approximate $p(x_0:k \mid z_1:k)$ using MCMC. First an initial state $x_0$ is sampled according to a proposed density, which can be Gaussian and where the mean and covariance are computed from the Monte Carlo samples of the last time step. Subsequent state vectors are then sampled through the state function step by step. The target density used in the accept/reject step is the non-normalized posterior density:

$$p(x_0:k \mid z_1:k) \propto p(x_0)\prod_{i=1}^{k} p(z_i \mid x_i)$$

As the computational cost of MCMC increases with time, MCMC is performed during the first few time steps, before switching over to the conventional particle filter. Experiments have shown that an MCMC running for only 10 steps can significantly improve state estimation performance whilst incurring reasonable computational load.
4. Simulation Results

The performance of particle filters is demonstrated through a simulated exothermic irreversible parallel reaction taking place in a continuous stirred tank reactor (CSTR). This example has been well studied in the literature with respect to state space estimation, (e.g. Valluri and Soroush, 1996 and references therein). There are two states in this system: $C_A$ (kmol · m$^{-3}$) is the concentration of the reactant, and $T$ (K) is the reactor temperature. In this study it is assumed that only $T$ can be measured on-line with Gaussian noise. Thus the objective of the state estimation task is to infer $C_A$ and $T$ sequentially given contaminated observations of $T$. The CSTR simulation (Valluri and Soroush, 1996) can exhibit multiple steady states. For example, when the heat input rate $q$ under steady state is $q_{ss} = -1.030$ kJ · s$^{-1}$, there are three steady-state operating points, as shown in Table 1. Of these steady-state operating points, SS$_2$ is unstable. In the simulation, two initial conditions, Table 1, are used to study filter performance under different steady states. The measurement signal-to-noise ratio was 20 db. One hundred simulations were realized to test the robustness of the particle filters, each running for 2000 secs with a sampling rate of 5 secs.

In the first experiment, it is assumed that the true initial conditions are known. The root mean square error (RMSE) is used to evaluate the performance of the filters. 95% confidence bounds were calculated for the RMSE based on one hundred realizations of the simulated process. Table 2 shows that the ASIR filter with only 50 particles results in a significantly lower estimation error than the EKF. The ASIR filter with an increased number of particles can bring further, although marginal improvements. The algorithms were implemented in C++ and run on a 3.06GHz Pentium-4 processor under a Windows XP operating system. The computing time of the ASIR filter is acceptable.

Table 1: Steady states for the CSTR with corresponding initial conditions.

<table>
<thead>
<tr>
<th>Steady States</th>
<th>$C_A$</th>
<th>$T$</th>
<th>Initial Conditions</th>
<th>$C_A$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS$_1$</td>
<td>8.0</td>
<td>310.0</td>
<td>IC$_1$</td>
<td>6.3</td>
<td>300.0</td>
</tr>
<tr>
<td>SS$_2$</td>
<td>3.3</td>
<td>370.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SS$_3$</td>
<td>1.3</td>
<td>400.0</td>
<td>IC$_3$</td>
<td>0.3</td>
<td>410.0</td>
</tr>
</tbody>
</table>

Table 2: State estimation error (RMSE) with known initial condition.

<table>
<thead>
<tr>
<th>Filters</th>
<th>Steady State 1</th>
<th>Steady State 3</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_A$</td>
<td>$T$</td>
<td>$C_A$</td>
</tr>
<tr>
<td>EKF</td>
<td>0.134 ± 0.016</td>
<td>0.905 ± 0.093</td>
<td>0.164 ± 0.014</td>
</tr>
<tr>
<td>ASIR(50)</td>
<td>0.040 ± 0.014</td>
<td>0.310 ± 0.088</td>
<td>0.031 ± 0.009</td>
</tr>
<tr>
<td>ASIR(500)</td>
<td>0.023 ± 0.006</td>
<td>0.301 ± 0.080</td>
<td>0.027 ± 0.008</td>
</tr>
</tbody>
</table>

Table 3: State estimation error (RMSE) with poor initial condition.

<table>
<thead>
<tr>
<th>Filters</th>
<th>Steady State 1</th>
<th>Steady State 3</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_A$</td>
<td>$T$</td>
<td>$C_A$</td>
</tr>
<tr>
<td>EKF</td>
<td>0.259 ± 0.011</td>
<td>1.416 ± 0.106</td>
<td>0.235 ± 0.013</td>
</tr>
<tr>
<td>ASIR(500)</td>
<td>0.206 ± 0.215</td>
<td>1.861 ± 0.997</td>
<td>0.179 ± 0.188</td>
</tr>
<tr>
<td>HASIR(500)</td>
<td>0.156 ± 0.606</td>
<td>1.372 ± 0.100</td>
<td>0.083 ± 0.037</td>
</tr>
<tr>
<td>MCMC</td>
<td>0.151 ± 0.020</td>
<td>0.369 ± 0.103</td>
<td>0.078 ± 0.022</td>
</tr>
</tbody>
</table>
When the initial conditions deviate significantly from the true values, the conventional ASIR filter converges slowly, leading to an overall high estimation error. Table 3 shows the results with poor initial conditions: $IC_1 = (5.8, 330)$ and $IC_3 = (0.6, 380)$. In this case, the ASIR filter, even with 500 particles, cannot outperform the EKF. However with the inclusion of a built-in EKF component, the hybrid ASIR (HASIR) filter can achieve improved estimation of concentration $C_A$ and a marginally better $T$. Overall the ASIR filter with the MCMC step provides good estimates of both states with acceptable computational cost. It is noted that the accurate recovery of $T$ from noisy measurement is also an important topic in data rectification (Singhal and Seborg, 2000).

5. Conclusions

This paper has introduced particle filtering for the estimation of a state space model. Particle filters are appropriate for handling general state space models and do not rely on the assumptions of linearity or a Gaussian posterior density. A particle filter algorithm is first described for the state estimation task. For poor initial conditions, a Markov chain Monte Carlo method was applied to enhance the estimation accuracy of the particle filters. The proposed method was then evaluated on a simulated CSTR, and promising results were obtained. Particle filters can be applied to existing techniques requiring the on-line estimation of state space models, such as model based quality monitoring, optimization, predictive control and data rectification. An extension to a large-scale polymerization process with uncertain kinetic parameters is on-going.

Acknowledgements

T. Chen would like to acknowledge the financial support from the EPSRC and the UK ORS Award for his PhD study.

References


Singhal, A. and Seborg, D., 2000, Dynamic data rectification using the expectation maximization algorithm, AIChE, 46(8), 1556.
