Merging particle filter for high-dimensional nonlinear problems

S. Nakano, G. Ueno, and T. Higuchi

(The Institute of Statistical Mathematics, ROIS, Tokyo, JAPAN)

1 Introduction

Sequential data assimilation is an on line approach in which a state at each time step is estimated as the observation is obtained. Most of sequential data assimilation techniques are based on Bayesian approach. The information of observations is incorporated into the system model by considering a posterior probability density function (PDF) derived from a prior PDF, where the prior PDF is obtained using past data and a system model. The particle filter (PF) (Gordon et al., 1993; Kitagawa, 1993, 1996) is one of such sequential data assimilation algorithms. The PF approximates PDFs at each time step by an ensemble of a large number of particles. An estimation of a posterior PDF is obtained by resampling with replacement from a prior ensemble. Since the PF does not require assumptions of linearity or Gaussianness, it is applicable to general nonlinear problems including cases with nonlinear observations which other algorithms such as the ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998) do not provide good estimation.

However, the PF often encounters a problem called 'degeneration', which does not occur in the EnKF. As resampling procedures are applied recursively, most of the particles are replaced by particles that fit the observed data better, and the posterior PDF is eventually represented by only a few of the particles among the members of the initial ensemble. This reduces the validity of ensemble approximation especially in applying to high-dimensional systems. Although this problem could be avoided by increasing the number of particles in the ensemble, the increase of the number of particles requires a prohibitive computational cost at each forecast step.

In order to overcome this problem, we have devised another technique, the merging particle filter (MPF). In the MPF, a filtering procedure is performed by merging several particles of a prior ensemble, which is rather similar to the genetic algorithm (e.g., Goldberg, 1989). This merging procedure is performed such that the first and second moments of a posterior PDF is preserved. The MPF allows the degeneration problem to be avoided without unduly increasing the number of particles and thus the MPF is applicable to higher-dimensional systems. Moreover, the primary advantage of the PF over the EnKF is inherited; that is, the MPF is applicable even to cases in which the relationship between a state and observed data is nonlinear. In our previous paper (Nakano et al., 2007), the basic idea of the MPF is given. The purpose of this paper is to briefly review the algorithm of the MPF and to discuss some properties of the MPF.

2 Particle filter

In the following, we consider a state space model as:

$$\boldsymbol{x}_k = F_k(\boldsymbol{x}_{k-1}, \boldsymbol{v}_k) \tag{1a}$$

$$\boldsymbol{y}_k = H_k(\boldsymbol{x}_k) + \boldsymbol{w}_k \tag{1b}$$

where the vectors $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$ indicate the state of a system and observed data at a discrete time $T = t_k$ $(k = 1, \ldots)$, respectively, and the vectors $\boldsymbol{v}_k$ and $\boldsymbol{w}_k$ denote system noise and observation noise, respectively. The operator F_k represents the temporal evolution from time t_{k-1} to time t_k according to the simulation. The operator H_k projects the state vector $\boldsymbol{x}_k$ to the observation space.

The PF considers a PDF of a state $\boldsymbol{x}_k$. The PDF is approximated by an ensemble consisting of a large number of discrete samples called 'particles'. Suppose that a filtered distribution at time $T = t_{k-1}$, $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$, is approximated by particles $\{\boldsymbol{x}^{(1)}_{k-1|k-1}, \boldsymbol{x}^{(2)}_{k-1|k-1}, \cdots, \boldsymbol{x}^{(N)}_{k-1|k-1}\}$ as

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta\left(\boldsymbol{x}_{k-1} - \boldsymbol{x}^{(i)}_{k-1|k-1}\right) \tag{2}$$

where δ is Dirac's delta function, and N is the number of particles in the ensemble. Here we expressed $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{k-1})$ as $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$. We obtain an ensemble approximation of the forecast distribution of the state at the next observation time $T = t_k$ as

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \boldsymbol{x}^{(i)}_{k|k-1}\right) \tag{3}$$

where $\boldsymbol{x}^{(i)}_{k|k-1}$ is given by $F_k(\boldsymbol{x}^{(i)}_{k-1|k-1}, \boldsymbol{v}^{(i)}_k)$ for each particle i. This procedure is called the forecast step.

From the forecast distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$ and observed data $\boldsymbol{y}_k$, we obtain an approximation of the filtered PDF $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$:

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) &= \frac{p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})\, p(\boldsymbol{y}_k|\boldsymbol{x}_k)}{\int p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})\, p(\boldsymbol{y}_k|\boldsymbol{x}_k) d\boldsymbol{x}_k} \\ &\approx \frac{1}{\sum_j p\left(\boldsymbol{y}_k|\boldsymbol{x}_{k|k-1}^{(j)}\right)} \sum_{i=1}^{N} p\left(\boldsymbol{y}_k|\boldsymbol{x}_{k|k-1}^{(i)}\right) \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k-1}^{(i)}\right) \\ &= \sum_{i=1}^{N} w_i \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k-1}^{(i)}\right) \end{aligned} \tag{4}$$

where $p(\boldsymbol{y}_k|\boldsymbol{x}_{k|k-1}^{(i)})$ is the likelihood of $\boldsymbol{x}_{k|k-1}^{(i)}$ given the data $\boldsymbol{y}_k$ and the weight w_i is defined as

$$w_i = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_{k|k-1}^{(i)})}{\sum_j p(\boldsymbol{y}_k|\boldsymbol{x}_{k|k-1}^{(j)})}. \tag{5}$$

This is called the filtering step.

We then obtain a new ensemble $\{\boldsymbol{x}_{k|k}^{(1)}, \cdots, \boldsymbol{x}_{k|k}^{(N)}\}$ by resampling the forecast ensemble $\{\boldsymbol{x}_{k|k-1}^{(1)}, \cdots, \boldsymbol{x}_{k|k-1}^{(N)}\}$ with replacement. The new ensemble may contain multiple duplicates originating from the same particle in the forecast ensemble. The resampling is performed such that the number of copies for a particle $\boldsymbol{x}_{k|k-1}^{(i)}$ is nearly proportional to the weight w_i. Denoting the number of copies for a particle $\boldsymbol{x}_{k|k-1}^{(i)}$ as m_i, m_i satisfies

$$m_i \approx N w_i \quad \left(\sum m_i = N;\ m_i \geq 0\right) \tag{6}$$

for each $\boldsymbol{x}_{k|k-1}^{(i)}$. Using Eq. (6), Eq. (4) can be further approximated as

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) &\approx \sum_{i=1}^{N} \frac{m_i}{N} \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k-1}^{(i)}\right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k}^{(i)}\right). \end{aligned} \tag{7}$$

Thus, the newly generated ensemble approximates the filtered PDF $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. Equation (7) has the same form as Eq. (2). We can then recursively repeat the above procedure from Eq. (2) to Eq. (7) to incorporate a sequence of observed data into the system model.

3 Merging particle filter

The MPF is a variant of the PF. In the MPF, each particle of a filtered ensemble is generated by combining multiple particles drawn from the forecast ensemble, which is rather similar to the genetic algorithm. When the number of particles to be combined is denoted by n, it is necessary to draw $n \times N$ samples from the forecast ensemble. to obtain an ensemble: $\{\hat{\boldsymbol{x}}_{k|k}^{(1,1)}, \cdots, \hat{\boldsymbol{x}}_{k|k}^{(n,1)}, \cdots, \hat{\boldsymbol{x}}_{k|k}^{(1,N)}, \cdots, \hat{\boldsymbol{x}}_{k|k}^{(n,N)}\}$. The sampling of the $n \times N$ samples is performed such that a particle $\boldsymbol{x}_{k|k-1}^{(i)}$ is drawn with a probability w_i like the PF described in the previous section. It should be noted that a subset $\{\hat{\boldsymbol{x}}_{k|k}^{(j,1)}, \cdots, \hat{\boldsymbol{x}}_{k|k}^{(j,N)}\}$ affords an ensemble approximation of the filtered PDF as

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^{(j,i)}\right) \tag{8}$$

because copies of each particle $\boldsymbol{x}_{k|k-1}^{(i)}$ is contained with a ratio proportional to w_i in the subset. We then make a new ensemble consisting of N particles $\{\boldsymbol{x}_{k|k}^{(1)}, \cdots, \boldsymbol{x}_{k|k}^{(N)}\}$. Each particle in this new ensemble is generated as a weighted sum of n samples from the $n \times N$ sample set as:

$$\boldsymbol{x}_{k|k}^{(i)} = \sum_{j=1}^{n} \alpha_j \hat{\boldsymbol{x}}_{k|k}^{(j,i)}. \tag{9}$$

In order that the newly generated ensemble approximately preserves the mean and covariances of the filtered PDF, the weights α_j are set to satisfy

$$\sum_{j=1}^{n} \alpha_j = 1 \tag{10a}$$

$$\sum_{j=1}^{n} \alpha_j^2 = 1 \tag{10b}$$

where each α_j is a real number. When the merging weights satisfy Eq. (10a), the mean of the PDF approximated by the new ensemble $\{\boldsymbol{x}_{k|k}^{(1)}, \cdots, \boldsymbol{x}_{k|k}^{(N)}\}$ becomes

$$\begin{aligned}
\int \boldsymbol{x}_k \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k}^{(i)}\right)\, d\boldsymbol{x}_k &= \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_{k|k}^{(i)} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n} \alpha_j \hat{\boldsymbol{x}}_{k|k}^{(j,i)} \\
&= \sum_{j=1}^{n}\left[\alpha_j \int \boldsymbol{x}_k \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^{(j,i)}\right)\, d\boldsymbol{x}_k\right] \\
&\approx \sum_{j=1}^{n} \alpha_j \int \boldsymbol{x}_k\, p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})\, d\boldsymbol{x}_k \\
&= \int \boldsymbol{x}_k\, p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})\, d\boldsymbol{x}_k = \boldsymbol{\mu}_{k|k}
\end{aligned} \tag{11}$$

where $\boldsymbol{\mu}_{k|k}$ is the mean of the filtered PDF $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. In addition, if the merging weights α_j satisfy Eq. (10b), the covariances given by the new ensemble become

$$\begin{aligned}
&\int (\boldsymbol{x}_k - \boldsymbol{\mu}_{k|k})(\boldsymbol{x}_k - \boldsymbol{\mu}_{k|k})^T \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k}^{(i)}\right) d\boldsymbol{x}_k \\
&\quad = \int \boldsymbol{x}_k \boldsymbol{x}_k^T \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k}^{(i)}\right)\, d\boldsymbol{x}_k - \boldsymbol{\mu}_{k|k}\boldsymbol{\mu}_{k|k}^T \\
&\quad = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_{k|k}^{(i)} \boldsymbol{x}_{k|k}^{(i)\,T} - \boldsymbol{\mu}_{k|k}\boldsymbol{\mu}_{k|k}^T \\
&\quad = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n} \alpha_j^2 \hat{\boldsymbol{x}}_{k|k}^{(j,i)} \hat{\boldsymbol{x}}_{k|k}^{(j,i)\,T} - \boldsymbol{\mu}_{k|k}\boldsymbol{\mu}_{k|k}^T \\
&\quad = \sum_{j=1}^{n} \alpha_j^2 \int \boldsymbol{x}_k \boldsymbol{x}_k^T \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^{(j,i)}\right)\, d\boldsymbol{x}_k - \boldsymbol{\mu}_{k|k}\boldsymbol{\mu}_{k|k}^T \\
&\quad \approx \int \boldsymbol{x}_k \boldsymbol{x}_k^T\, p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})\, d\boldsymbol{x}_k - \boldsymbol{\mu}_{k|k}\boldsymbol{\mu}_{k|k}^T = \boldsymbol{\Sigma}_{k|k}
\end{aligned} \tag{12}$$

where $\boldsymbol{\Sigma}_{k|k}$ is the covariance matrix of $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. Here, we used an approximation as

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{x}}_k^{(j_1,i)} - \boldsymbol{\mu}_{k|k})(\hat{\boldsymbol{x}}_k^{(j_2,i)} - \boldsymbol{\mu}_{k|k})^T \approx 0 \quad (\text{if } j_1 \neq \mathrm{J}_2),$$

which is justified because the two sets of samples $\{\hat{\boldsymbol{x}}_{k|k}^{(j_1,1)}, \cdots, \hat{\boldsymbol{x}}_{k|k}^{(j_1,N)}\}$ and $\{\hat{\boldsymbol{x}}_{k|k}^{(j_2,1)}, \cdots, \hat{\boldsymbol{x}}_{k|k}^{(j_2,N)}\}$ are obtained through independent random sampling and would not correlate with each other. The ensemble obtained using Eq. (9) therefore gives an approximation of $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ preserving the mean and covariances as

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \approx \frac{1}{N}\sum_{i=1}^{N} \delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k|k}^{(i)}\right). \tag{13}$$

4 Discussion

The MPF approximately preserves the mean and covariance (i.e., the first and second moments) of the filtered PDF, and it asymptotically preserves the mean and covariance as the number of particles approaches infinity. However, the moments of higher order than the second moment are not preserved and thus the MPF does not ensure that the shape of the filtered PDF is preserved. It should be noted that the moments of higher order than the second moment are not preserved even if the number of merged particles n is increased. Here we assume that the dimension of x_k is one for simplicity. The m-th moment of x_k is then given as

$$\begin{aligned}
&\int (x_k - \mu_{k|k})^m \, p(x_k|y_{1:k}) \, dx_k \\
&\approx \int (x_k - \mu_{k|k})^m \frac{1}{N} \sum_{i=1}^{N} \delta \left(x_k - x_{k|k}^{(i)} \right) dx_k \\
&= \frac{1}{N} \sum_{i=1}^{N} (x_{k|k}^{(i)} - \mu_{k|k})^m = \frac{1}{N} \sum_{i=1}^{N} \prod_{l=1}^{m} \left(\sum_{j_l=1}^{n} \alpha_{j_l} \hat{x}_{k|k}^{(j_l,i)} - \mu_{k|k} \right) \\
&\approx \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} \alpha_j^m \left(\hat{x}_{k|k}^{(j,i)} - \mu_{k|k} \right)^m = \sum_{j=1}^{n} \alpha_j^m \int (x_k - \mu_{k|k})^m \frac{1}{N} \sum_{i=1}^{N} \delta \left(x_k - \hat{x}_{k|k}^{(j,i)} \right) \, dx_k \\
&\approx \sum_{j=1}^{n} \alpha_j^m \int (x_k - \mu_{k|k})^m \, p(x_k|y_{1:k}) \, dx_k
\end{aligned} \tag{14}$$

where we use an approximation as

$$\frac{1}{N} \sum_{i=1}^{N} \prod_{l=1}^{m} \left(\hat{x}_{k|k}^{(j_l,i)} - \mu_{k|k} \right) \approx 0 \quad (\text{unless } j_1 = j_2 = \cdots = j_m).$$

In order to preserve the m-th moment even after taking the weighted sum as Eq. (9), a set of weights $\{\alpha_j\}_{j=1}^n$ must satisfy

$$\sum_{j=1}^{n} \alpha_j^m = 1. \tag{15}$$

On the other hand, if the set of real numbers $\{\alpha_j\}_{j=1}^n$ satisfy Eq. (10b), $0 < |\alpha_j| < 1$ are satisfied for all j and thus

$$|\alpha_j|^2 > |\alpha_j|^m \tag{16}$$

if an integer m is larger than 2. Since Eq. (16) is satisfied for all j,

$$\sum_{j=1}^{n} \alpha_j^2 > \sum_{j=1}^{n} \alpha_j^m. \tag{17}$$

If Eq. (10b) is satisfied and $m > 2$,

$$\sum_{j=1}^{n} \alpha_j^m < 1. \tag{18}$$

Therefore, if Eq. (10b) is satisfied, Eq. (15) can not be satisfied when $m > 2$. This means that there does not exist a set of weights $\{\alpha_j\}_{j=1}^{n}$ which preserves the second moment and moments of higher order than the second moment simultaneously.

In implementing the MPF, the number of merged particles n can be chosen almost arbitrarily. However, n must be equal to or greater than 3. If $n = 1$, the weight α_1 must be 1 in order to satisfy both Eq. (10a) and Eq. (10b), which is obviously equivalent to the normal PF. If $n = 2$, then one of merging weights must be 1, and the other must be 0, so as to satisfy both Eq. (10a) and Eq. (10b). This setting is also equivalent to the normal PF, which means that the merging procedure does not make sense. Although there is no upper limit for n, it is not necessary to set n to be large.

Although the filtered ensemble in the MPF may contain duplicate particles, the number of the duplicate particles is much smaller than that for the PF. If the weights are set such that no two weights are equal to each other and that none of the weights are zero, the number of the duplicate particles can be remarkably reduced in comparison with the PF. In the MPF, two duplicate particles in the filtered ensemble $\{\boldsymbol{x}_{k|k}^{(1)}, \cdots, \boldsymbol{x}_{k|k}^{(N)}\}$ can be generated only from two identical sets of n particles in the forecast ensemble, if duplicate particles are not contained in the forecast ensemble $\{\boldsymbol{x}_{k|k-1}^{(1)}, \cdots, \boldsymbol{x}_{k|k-1}^{(N)}\}$. When the probability that a particle $\boldsymbol{x}_{k|k-1}^{(i)}$ is drawn from the forecast ensemble is w_i ($0 \leq w_i < 1$), the probability that a sequence of n particles $\{\boldsymbol{x}_{k|k-1}^{(i_1)}, \cdots, \boldsymbol{x}_{k|k-1}^{(i_n)}\}$ is drawn is $\prod_{j=1}^{n} w_{i_j}$. Thus, the number of duplicate particles contained in the filtered ensemble is, at most, approximately $N \times (\max w_i)^n$ for the MPF, while it is $N \times \max w_i$ for the PF.

When n is equal to or greater than 3, there are infinite allowable sets of merging weights: $\{\alpha_1, \cdots, \alpha_n\}$. Although there is no definitive way to determine the values of the weights, it would be better to set them such that no two weights are equal to each other and that none of the weights are zero in order to maintain the diversity of the ensemble. However, characteristics of an ensemble obtained as a result of the MPF depend to some extent on how the set of merging weights is given. For example, if one of the weights is nearly equal to 1 and the other weights are nearly equal to zero, the MPF shows a similar behavior to the PF; that is, the shape of a filtered PDF is mostly preserved but the degeneration problem tend to be serious (Nakano et al., 2008).

5 Summary

The merging particle filter (MPF) provides an ensemble approximation of the filtered PDF. Each member of a filtered ensemble is generated from a weighted sum of multiple samples from the forecast ensemble such that the mean and covariance of the filtered distribution are approximately preserved. The MPF can not preserve the shape of the filtered PDF because the moments of higher order than the second moment are not preserved. However, when one merging weights is set to be near 1 and the other weights are set to be small, it shows s similar behavior to the PF; that is, the shape of a filtered PDF is mostly preserved although the degeneration problem could be serious.

References

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998., Analysis scheme in the ensemble Kalman filter, *Mon. Wea. Rev.*, *126*, 1719.

Evensen, G., 1994., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99(C5)*, 10143.

Goldberg, D. E., Genetic algorithms in search, optimization and machine learning, Addison-Wesley, Reading, 1989.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993., Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F*, *140*, 107.

Kitagawa, G., 1993., Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space model, *Inst. Statist. Math. Res. Memo.*.

Kitagawa, G., 1996., Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *J. Comp. Graph. Statist.*, *5*, 1.

Nakano, S., G. Ueno, and T. Higuchi, 2007., Merging particle filter for sequential data assimilation, *Nonlin. Process. Geophys.*, *14*, 395–408.

Nakano, S., G. Ueno, K. Nakamura, and T. Higuchi, 2008., The merging particle filter and its characteristics (in Japanese), *Proc. of Inst. of Statist. Math.*, *56*, 225–234.