Particle Kalman Filtering for Data Assimilation in Meteorology and Oceanography

Ibrahim Hoteit[1], Din-Tuan Pham[2], George Triantafyllou[3], Gerasimo Korres[3]

[1]Physical Oceanography Research Division, Scripps Institution of Oceanography, USA
[2]Centre National de la Recherche Scientifique, Université de Grenoble, France
[3]Institute of Oceanography, Hellenic Center of Marine Research, Greece
Correspondence: ihoteit@ucsd.edu

INTRODUCTION

We describe a discrete solution of the optimal nonlinear filter that generalizes the optimality of the correction step of the ensemble Kalman filters to nonlinear systems. This approach is based on a Gaussian mixture representation of the state probability density function which results in a new particle-type filter, called particle Kalman filter (PKF), in which the standard (weight-type) particle filter correction is complemented by a Kalman-type correction for each particle using the associated covariance matrix in the Gaussian mixture. The optimal solution of the nonlinear filtering problem is then obtained as the weighted average of an ensemble of Kalman filters operating in parallel. The Kalman-type correction reduces the risk of ensemble collapse, which enables the filter to efficiently operate with fewer particles than the particle filter. Running an "ensemble of Kalman filters" is however computationally prohibitive for high dimensional systems. We first derive the popular ensemble Kalman filter as a suboptimal variant of the PKF and evaluate its performances against the PKF with the Lorenz-96 model. Then we discuss different approaches to reduce the computational burden of the PKF filter for application to atmospheric and oceanic data assimilation problems.

THE OPTIMAL NONLINEAR FILTERING

Consider the nonlinear stochastic discrete-time dynamical system

$$\begin{cases} x_k = M_k(x_{k-1}) + \eta_k \\ y_k = H_k(x_k) + \varepsilon_k \end{cases}$$

where at time t_k, x_k is the state vector, y_k is the observation vector, M_k and H_k are two continuously differentiable operators respectively representing the transition operator (model) and the observational operator. η_k and ε_k denote the dynamical and the observational noise which we assume independent Gaussian variables with mean zero and covariance matrices Q_k and R_k, respectively.

Starting from a random initial condition x_0 with a known probability density function (*pdf*), the optimal nonlinear filter provides the conditional density function $p_k(\cdot \mid y_{1:k})$ of the system state given all available measurements up to the estimation time $y_{1:k} = (y_1, \ldots, y_k)$ in two steps which we summarize below. The reader is referred to (Doucet, 2001) for an extensive description of the filter. With the state *pdf* in hand, one can determine different estimates/statistics of the system state, as the minimum variance estimate (MV), or the maximum likelihood estimate (ML).

- *Forecast step*: The model is integrated forward starting from the available analysis density $p_{k-1}(\cdot \mid y_{1:k-1})$ at time t_{k-1} to compute the predictive density $p_{k|k-1}(\cdot \mid y_{1:k-1})$ at the time of the next available observation t_k via the Chapman-Kolmogrov equation

$$p_{k|k-1}(x_k \mid y_{1:k-1}) = \int p(x_k = x \mid x_{k-1} = u) p_{k-1}(u \mid y_{1:k-1}) du$$

where $p(x_k = x \mid x_{k-1} = u) = \phi\left(x - M_k(u), Q_k\right)$ is the conditional density of the state vector x_k to be at x at time t_k given that it was at u at time t_{k-1}, where $\phi\left(z;\Sigma\right) = 1/\sqrt{\det(2\pi\Sigma)}\exp(-z^T\Sigma^{-1}z/2)$ denotes the Gaussian density of zero mean and covariance matrix Σ.

- *Correction step*: When the new observation y_k is available, update the analysis density with the Bayes rule,

$$p_k(x \mid y_{1:k}) = \frac{1}{b_k} p_{k|k-1}(x \mid y_{1:k-1}) \phi\left(y_k - H_k(x); R_k\right)$$

The analysis density is therefore obtained by multiplying the prior predictive density by the observation likelihood and normalizing by a constant $b_k = \int p_{k|k-1}(u \mid y_{1:k-1}) \phi\left(y_k - H_k(u); R_k\right)$ to ensure a probability density.

KALMAN, ENSEMBLE, AND PARTICLE FILTERS

A simple form of filtering is the Kalman filter (KF) which provides the MV estimate of the state for linear systems with Gaussian error statistics. Under these conditions, the *pdf*s of the system state given the available observations are always Gaussian. They are thus determined by their mean and covariance matrix. The KF allows to recursively compute these parameters as follows:

- *Forecast step*: Starting from the analysis state x_{k-1}^a at time t_{k-1} and the associated error covariance matrix P_{k-1}^a, the model is integrated forward in time to compute the forecast state as $x_k^f = M_k(x_{k-1}^a)$, and to update the error covariance matrix with $P_k^f = M_k P_{k-1}^a M_k^T + Q_k$.
- *Correction step*: Correct the forecast with the new observation using the standard Kalman correction equation

$$x_k^a = x_k^f + G_k\left(y_k - H_k(x_k^f)\right),$$

where $G_k = P_k^f H_k^T \Sigma_k^{-1}$ is the called Kalman gain and $\Sigma_k = H_k P_k^f H_k^T + R_k$ is the innovation matrix. The associated error covariance matrix is then $P_k^a = (I - G_k H_k) P_k^f$.

The KF is only designed for linear systems. Since most dynamical and observational systems encountered in practice are nonlinear, the system equations were often linearized about the most recent estimate, leading to the popular, but no longer optimal, Extended Kalman filter (EKF). The EKF algorithm is the same as the KF except for the use of the linearized operators of M_k and H_k in the update equations of the error covariance matrices. Several studies, e.g. Gauhier et al. (1992), have however demonstrated that the linearization of the system produces instabilities, even divergence, when applied to strongly nonlinear systems.

A fully nonlinear discrete solution of the optimal nonlinear filter problem can be obtained from the particle filter (PF) which uses point mass representation $\sum w^i \delta_{x^i}$, also called mixture of Dirac distributions, of the state *pdf*s (Doucet, 2001). In this filter, each particle x^i is assigned a weight w^i that is updated by the filter correction step. The MV state estimate is then the weighted-mean of the particles ensemble $\sum w^i x^i$. Starting from an ensemble of particles, the parameters x^i and w^i are updated as follows:

- *Forecast step*: The particles x_{k-1}^{i} are integrated forward with the model to the time of the next available observation y_k.
- *Correction step*: Use the new information from the observation to update the weights according to

$$w_k^i = \frac{1}{c_k} w_{k-1}^i \phi\left(y_k - H_k(x_k^i); R_k\right),$$

  where c_k is a constant normalizing the total weight. The particles x_k^i remain unchanged after the correction step. A particle will receive more weight proportional to how close this particle is from the new observation. The misfit between a particle and the observation is measured with respect to the observational error covariance matrix R_k.

In practice, the PF suffers from a major problem known as the degeneracy phenomenon (Doucet et al., 2001); after several iterations most weights becomes concentrated on very few particles which means that only a tiny fraction of the ensemble contributes to the ensemble mean. This happens because most of the particles drift away from the true state since they are not updated by the observations. This leads very often to the divergence of the filter. The use of more particles could only alleviate this problem over short time periods, and the only possible way to get around it is "resampling" (Doucet, 2001). This technique basically consists of drawing new particles according to the distribution of the ensemble and then reassigning them the same weights. Besides being computationally intensive, this approach introduces Monte Carlo fluctuations which can seriously degrade the filter's performance. In practice, even with resampling, the filter still requires a large number of particles to provide acceptable performances. This makes brute-force implementation of the PF problematic with high dimensional systems. Interesting discussions about the potential of the optimal nonlinear filter for atmospheric and oceanic data assimilation systems can be found in Anderson and Anderson (1999), Kivman (2003) and Van Leeuwen (2003).

To avoid the problems of the PF, Evensen (1994) introduced the popular Ensemble Kalman filter (EnKF) as a hybrid approach between the KF and the PF. More precisely, the EnKF also makes use of an ensemble of particles. It has the same forecast step as the PF, but not the same correction step. The EnK filter actually retains the "linearity aspect" of the Kalman filter in the analysis, in that it applies the Kalman correction to all particles using forecast error covariance matrices estimated from the particles ensemble. The weights are kept uniform. The EnKF can be summarized as follows:

- *Forecast step*: As in the PF, the "analyzed" particles $x_{k-1}^{a,i}$ are integrated forward with the model to determine the "forecast" particles $x_k^{f,i}$ at time t_k.
- *Correction step*: Use the new observation to correct the forecast particles with the KF correction step,

$$x_k^{a,i} = x_k^{f,i} + G_k\left(y_k^i - H_k(x_k^{f,i})\right).$$

Here, the Kalman gain G_k is computed from the sample covariance matrix of the forecast particles which is assumed to represent the uncertainty around the forecast state. Note that the observation has a superscript index giving the particle number in the correction equation of the particles. This is because the observations need to be perturbed by noise generated from the *pdf* of the observational error so that the sampled covariance matrix of the analysis ensemble matches the analysis error covariance matrix of the KF (Burger et al., 1999). The uncertainties on the state estimates are therefore described by an ensemble of particles and their covariance matrices are not required for the filter's algorithm.

The correction step of the Kalman filter uses only the first two moments of the particle ensemble, and is thus suboptimal for non-Gaussian systems. In practical situations however, the EnKF was found to be more robust than the PF when small-size ensembles were used thanks to the Kalman update of its particles which significantly reduces the risk of ensemble degeneracy by pulling the particles toward the true state of the system (Kivman, 2003; Van Leeuwen, 2003). However, the PF always outperformed the EnKF when large enough ensembles were used, with the EnKF showing insignificant improvements with increasing numbers of particles (Nakano et al., 2007).

THE PARTICLE KALMAN FILTER

The idea behind the particle Kalman filter (PKF) is to represent the state *pdf*s in the optimal nonlinear filter by mixture of Gaussian densities $\sum w^i \varphi\left(\cdot - x^i; P^i\right)$ centered around an ensemble of "particles" x^i each associated with a covariance matrix P^i. A Gaussian mixture has been already used by Anderson and Anderson (1999), Chen and Liu (2000) and Bengtsson et al. (2003) in the context of the nonlinear filter. It is expected to provide more reliable representation of the state *pdf*s than a simple mixture of Dirac functions used in the PF. Using this representation of the state *pdf*s in the optimal nonlinear filter and under the assumption that the covariance matrices of the mixture P^i are small, the parameters w^i, x^i, P^i are recursively updated as follows:

- _Forecast step_: Integrate the analysis particles $x_{k-1}^{a,i}$ and the associated covariance matrices in the mixture $P_{k-1}^{a,i}$ with the forecast step of the EKF.
- _Correction step_: Use the new observation to correct the forecast particles with the EKF correction step,

  $$x_k^{a,i} = x_k^{f,i} + G_k^i\left(y_k - H_k(x_k^{f,i})\right),$$

  where the Kalman gain G_k^i is now function of $P_k^{f,i}$, and to update the weights as in the PF

  $$w_k^i = \tfrac{1}{c_k} w_{k-1}^i \phi\left(y_k - H_k(x_k^i); \Sigma_k^i\right).$$

The PKF applies two corrections; a "Kalman-type correction" to the particles as in the EnKF but using the covariance matrix of the mixture instead of the sample covariance matrix of the ensemble, and a "particle-type correction" to the weights as in the PF but using the innovation matrix Σ_k^i associated with each particle $x_k^{f,i}$ instead of the observational error covariance matrix. The PKF is basically a "weighted ensemble of EKFs" running simultaneously. Note that a resampling step might also be needed in the PKF, but not as often as in the PF. This is because the Kalman-type correction reduces the degeneracy of the particles as in the EnKF, and the weights are distributed more uniformly than in the PF because the particles/data misfits are evaluated with respect to the innovation matrix Σ_k^i which is always greater than the observational error covariance matrix R_k. Our resampling scheme from a Gaussian mixture *pdf* is based on the Kernel density estimator method (Silverman, 1986). More details can be found in Hoteit et al. (2008).

PKF vs. EnKF: APPLICATION TO LORENZ-96

One can easily see that the EnKF is a sub-optimal PKF. The EnKF algorithm can be derived from the PKF algorithm as follows:

- Use uniform weights.
- Use the sample covariance matrix to correct the particles. But in this case, the observations need to be perturbed in order to match the KF correction.

We implemented the EnKF, the PKF and another variant of the PKF in which the weights were kept uniform with the strongly nonlinear Lorenz-96 model (Lorenz, 1996). The Lorenz-96 model mimics the time-evolution of a scalar atmospheric quantity and is described by the following set of equations:

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + f \qquad \text{for} \quad j = 1,\ldots,J$$

The nonlinear quadratic terms simulate advection, the linear term represents dissipation, and f is a constant external forcing. In its current configuration, $J = 40$ variables and boundary conditions are cyclic, i.e. $x_{-1} = x_{N-1}$, $x_0 = x_{N-1}$ and $x_{N+1} = x_1$. This model behaves chaotically in the case of external forcing $f = 8$. The model was discretised using Runge-Kutta fourth order scheme with a time step $\Delta t = 0.05$, which corresponds to 6 hours in real time. After a spin up integration, the model was integrated for one year to sample a set of reference states. Observations were extracted from the reference states after adding random Gaussian errors of mean zero and standard deviation 1. One out of two variables was assumed to be observed every day (4 model time steps). The root mean square (*rms*) misfit between the reference states and the filters estimates averaged over the one year assimilation period was used to evaluate the filters performances.

Ensemble Size / Filter	**EnKF**	**PKF**	**PKF - now**
N = 50	1.4	0.79	0.85
N = 100	0.87	0.73	0.76
N = 250	0.75	0.65	0.71

Table 1: *rms misfits between the reference states and the filters estimates averaged over the one year assimilation period as they result from the ensemble Kalman filter (EnKF), the particle Kalman filter (PKF), and the particle Kalman filter with uniform weights (PKF-now) using different number of particles* N = 50, 100, 250.

The averaged *rms* misfits between the reference states and the estimated states are presented in Table 1 for the EnKF, the PKF, and the PKF with uniform weights (no weights update – PKF-now) for three different ensemble sizes (N = 50, 100, and 250). In all cases, the PKF outperforms the EnKF with or without update of the weights. The comparison between the results of PKF and PKF-now suggests that updating the weights is important to determine the most accurate estimates. The good performance of the EnKF shows that the sample covariance matrix of the ensemble provides a good approximation for the covariance matrices of the mixture to correct the particles by the new observation. However, the better results obtained by the PKF-now compared with those of the EnKF clearly demonstrates that some information were lost by this approximation. Similar results were obtained from several experiments with different observations strategies. The main conclusion from these experiments is that the more we simplify the PKF algorithm, the less accurate is the final solution. It is not clear, however, what type of information were lost after these simplifications (i.e. no weights update and the use of the sample covariance matrix for the particles update) and more advanced studies should be performed with different setups and different models to analyze in depth the behavior of the different filters.

DISCUSSION

The PKF is a discrete solution of the optimal nonlinear filter based on a Gaussian mixture representation of the state *pdf*. It sets a new framework for the EnKF methods as it generalizes the optimality of their analysis step to nonlinear systems. Preliminary results with the simple Lorenz-96 model suggest that the PKF outperforms the EnKF even with small number of particles. This is because the PKF complements the EnKF-like correction of the particles by a PF-like correction of the weights. It further uses the covariance matrix of the Gaussian mixture, as derived from the optimal nonlinear filtering theory, to correct the particles when the EnKF uses the sample covariance matrix of the particles, which is only "suboptimal".

The PKF simultaneously runs a weighted ensemble of EKFs. This is far beyond our computing capabilities when dealing with computationally demanding atmospheric and oceanic models. To implement the PKF, or a variant of it, with such highly dimensional models, one must consider a low-rank parameterization of the Gaussian mixtures covariance matrices of the state *pdf*s. This is a very common approach in the atmospheric and oceanic Kalman filtering community (Hoteit et al., 2002). A low-rank PKF was already introduced by Hoteit et al. (2008). One can also implement the PKF by running a weighted ensemble of EnSRFs (Ensemble Square Root filters). This would generalize Houtekamer and Mitchell (1998)'s work who used a pair of EnKFs to deal with the problem of inbreeding. Other variants can be also derived from the PKF by considering different parameterizations and/or applying different simplifications (Hoteit et al., 2002) of the PKF mixture covariance matrices. Another direction for improvement would be to work on the correction step of the particles weights. The final goal is to develop a set of computationally feasible "suboptimal" PKFs that can outperform EnKF methods. As stated by Anderson (2003), developing filters in the context of the optimal nonlinear filtering problem, rather than starting from the Kalman filter, can lead to a more straightforward understanding of their capabilities.

REFERENCES

Anderson, J., 2003: A Local Least Squares Framework for Ensemble Filtering. Mon. Wea. Rev., **131**, 634-642.

Anderson, J., and Anderson S., 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Wea. Rev., **127**, 2741-2758.

Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. J. Geophys. Res., **108**, 8775, doi:10.1029/2002JD002900.

Chen, R., and J. Liu, 2000: Mixture Kalman filters. J. Roy. Statist. Soc., **62**, 493-508.

Doucet, A., N. de Freitas, and N. Gordon (2001): Sequential Monte Carlo methods in practice. Springer, New York, pp.581.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res., **99**, 10143-10162.

Gauthier, P., P. Courtier, and P. Moll, 1993: Assimilation of simulated wind Lidar data with a Kalman filter. Mon. Wea. Rev., **121**, 1803-1820.

Hoteit, I., D.-T. Pham, G. Triantafyllou, and G. Korres, 2008: A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. Mon. Wea. Rev., **136**, 317-334.

Hoteit, I., D.-T. Pham and J. Blum: Simplified reduced-order Kalman filtering and application to altimetric data assimilation in the tropical Pacific. J. Mar. Sys., **236**, 101-127, 2002.

Houtekamer, P. L., and L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev., **126,** 796–811.

Lorenz, E. N., 1996: Predictability: A problem partially solved, Proceedings of the Seminar on Predictability, vol. 1, pp. 1 – 18, Eur. Cent. for Medium- Range Weather Forecasts, Reading Berkshire, U.K., 1996.

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

Silverman, B.W., 1986: Density Estimation for Statistics and Data Analysis. *Chapman and Hall, 175pp.*

Van Leeuwen, P. J., 2003: A variance-minimizing filter for large-scale applications. Mon. Wea. Rev., **131**, 2071-2084.