Ensemble Kalman Filter

I. Kasanický
Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic,
Institute of Computer Science, Academy of Sciences of the Czech Republic.

K. Eben
Institute of Computer Science, Academy of Sciences of the Czech Republic.

Abstract. The ensemble Kalman filter (EnFK) has recently become one of the most popular methods for high dimensional data assimilation and it is widely used in many disciplines such as discretization of partial differential equations in geophysics or image reconstruction. The EnKF has been originally proposed as a Monte Carlo approximation to the extended Kalman filter, which is in practice inapplicable when dimension of the input data is huge.
Numerous publications have studied applications and asymptotic results of the EnKF, while properties of the Kalman filter (KF) and EnKF for the infinite dimensional Hilbert spaces are still under the development.
The main purpose of this paper is a brief description of the high dimensional EnKF with references to the most valuable publications. Also, a short introduction to problems related with KF and EnKF on infinite dimensional Hilbert spaces is included.

Introduction

The ensemble Kalman filter (EnKF) is a sophisticated sequential data assimilation method, used especially for high dimensional data. It is a Monte Carlo approximation of the Kalman filter (KF), where the true covariance matrix in the KF is replaced by the sample covariance matrix computed from the ensemble. Therefore it could be implemented very efficiently. This method was published for the first time in Evensen [1994].

Data assimilation is a statistical method for estimating the true state of a system (typically dynamic system evolving in time) by merging various measurements irregularly distributed in space and time, with a prior knowledge of the state. It has been widely used in many disciplines such as image reconstruction, weather prediction or wildfire modelling.

The paper is structured as follows. The first section is meant as a brief introduction to data assimilation problem, it presents necessary assumptions needed and ends with definition of the sequential estimation formula. In the next section the KF’s equations are reminded. The third section is used to derive the EnKF from the KF with references to relevant papers. The last section presents a summary of the main questions, which arise when we try to use KF on general Hilbert space.

For readers interested in implementation of the EnKF we recommend papers Evensen [2003] and Mandel [2009], where a discussion about efficient implementation of EnKF is included. Very good, but also very extensive, is the book Evensen [2009]. Many scripts, usually written in Matlab, with further discussion and development could be found on the website http://enkf.nersc.no/, which is administrated directly by Geir Evensen.

Preliminaries

Suppose we want to predict temperature in Europe. In that case we would probably use some geophysical model and make prediction of the temperatures in every point in some three-dimensional grid covering Europe, say, for the next day. The quality of the prediction depends
very much on the error of initial conditions which the model starts from. Usually we have some prior estimate of initial conditions, e.g. from a previous run of the model. The initial conditions can be improved with the help of available measurements.

It is obvious that it is not possible to have real measurements in all points of the grid, so the dimension of the measured data would be much lower than dimension of our model state. Thus we come to an estimation problem, which in geophysical sciences is called data assimilation. This terminology reflects the dominant role of the model which has to keep physical balance and the perturbation caused by statistical estimation procedure needs to be limited in practice.

So we assume that we have available data in some discrete time steps \( t_1, \ldots, t_K \) and will denote \( D_{t_i}, D_{t_i} \in \mathbb{R}^m \) data observed at the time \( t_i \). Similarly we will denote \( X_{t_i} \) a vector of size \( n \), which contains all the values describing our model state at the time \( t_i \). The dependence between \( D_{t_i} \) and \( X_{t_i} \) is characterized by the function

\[
    h_{t_i} : X_{t_i} \rightarrow h_{t_i}(X_{t_i}) \sim D_{t_i},
\]

which is deterministic, but it could be different in the different time steps. This function is usually called observation operator. The length of \( D_{t_i} \) is often much lower then the length of \( X_{t_i} \), \( m \ll n \).

The model state is a random vector and we will assume, that its distribution has

- a bounded second moment and
- a density on \( \mathbb{R}^n \).

Model states are evolved in time using the known operator (function)

\[
    \mathcal{M} : (X_{t_i}, t_i, t_{i+1}) \rightarrow X_{t_{i+1}}.
\]

It could be for example numerical solution to some differential equations.

Our goal is to estimate the model state at some future time \( X_{t_K} \) using all available data until time \( t_{K-1} \). We will assume, that the model is Markovian of order 1

\[
    p(X_{t_K} | X_{t_1}, \ldots, X_{t_{K-1}}) = p(X_{t_K} | X_{t_{K-1}}),
\]

where \( p(\cdot) \) denotes density function. The main theoretical background of data assimilation methods is the Bayes theorem, which states that

\[
    p(X_{t_K}^a) = p(X_{t_K}^f | D_{t_K}) \propto p(D_{t_K} | X_{t_K}^f) p(X_{t_K}^f),
\]

where \( X_{t_K}^f \) represents the forecast model state, in Bayesian statistics called also prior state, and \( X_{t_K}^a \) represents posterior model state, often called "analysis" by the meteorologists. Using Bayesian rule (2) together with Markov characteristics (1) we obtain the sequential estimation formula

\[
    p(X_{t_K}^a) \propto p(D_{t_K} | X_{t_K}^f) p(X_{t_K}^f | X_{t_1}, \ldots, D_{t_{K-1}}).
\]

### The Kalman Filter

One solution to the problem describe above is well known for almost 50 years. It was first proposed by Kalman [1960] and Kalman and Bucy [1961] and is known as Kalman Filter.

The KF restricts model function to be linear, so it could be rewritten using an \( n \times n \) matrix \( M_{t_i} \) and \( n \)-dimensional vector \( b_{t_i} \) in the form

\[
    \mathcal{M}(X_{t_i}, t_i, t_{i+1}) = M_{t_i} X_{t_i} + b_{t_i}.
\]
The linearity of $\mathcal{M}$ is quite restrictive and in many application the nonlinear operator $\mathcal{M}$ is replaced by linear approximation. In this case the matrix $\mathbf{M}_{ti}$ is replaced by Jacobian of the model $\mathcal{M}$ evaluated at time $t_i$ and the vector $\mathbf{b}_{ti}$ is replaced by the model evaluated at zero, that means by $\mathcal{M}(\mathbf{0}, t_i, t_{i+1})$. The KF also requires observation function to be linear and then it could be rewritten in similar matrix form

$$h_{ti}(\mathbf{X}_{ti}) = \mathbf{H}_{ti}\mathbf{X}_{ti} + \mathbf{h}_{ti},$$

where $\mathbf{H}_{ti}$ is $m \times n$ matrix and $\mathbf{h}_{ti}$ is vector of length $m$. Conditional distribution of the data available at time $t_i$ under condition of forecast model state is assumed to be normal

$$\mathbf{D}_{ti}\mathbf{X}_{ti}^f \sim N(\mathbf{H}_{ti}\mathbf{X}_{ti}^f, \mathbf{R}_{ti})$$

(4)

with some known non-singular covariance matrix $\mathbf{R}_{ti}$. The density of the distribution (4) is called data likelihood. Distribution of model state is also assumed to be normal

$$\mathbf{X}_{ti}^f \sim N(\mu_{ti}^f, \mathbf{Q}_{ti}^f),$$

(5)

where non-singular covariance matrix $\mathbf{Q}_{ti}^f$ is counted in each time step. We assume that $\mathbf{Q}_{ti}^f$ is known.

Under all these conditions, the distribution of analysis (posterior) model state remains normal. It can be shown that the posterior mean $\mu_{ti}^a$ and covariance matrix $\mathbf{Q}_{ti}^a$ are given by (6)–(8) written below:

$$\mathbf{K}_{ti} = \mathbf{Q}_{ti}^f(\mathbf{H}_{ti}\mathbf{Q}_{ti}^f\mathbf{H}_{ti}^\top + \mathbf{R}_{ti})^{-1},$$

(6)

$$\mu_{ti}^a = \mu_{ti}^f + \mathbf{K}_{ti}(\mathbf{D}_{ti} - \mathbf{H}_{ti}\mu_{ti}^f),$$

(7)

$$\mathbf{Q}_{ti}^a = (\mathbf{I} - \mathbf{K}_{ti}\mathbf{H}_{ti})\mathbf{Q}_{ti}^f.$$  

(8)

The matrix $\mathbf{K}_{ti}$ is called gain matrix and all three equations together are called Kalman formula update. So the KF could be summarized in a two step recursive algorithm:

- **analysis step**

$$\mu_{ti}^a = \mu_{ti}^f + \mathbf{K}_{ti}(\mathbf{D}_{ti} - \mathbf{H}_{ti}\mu_{ti}^f),$$

$$\mathbf{Q}_{ti}^a = (\mathbf{I} - \mathbf{K}_{ti}\mathbf{H}_{ti})\mathbf{Q}_{ti}^f.$$  

- **forecast step**

$$\mu_{ti+1}^f = \mathbf{M}_{ti}\mu_{ti}^a + \mathbf{b}_{ti},$$

$$\mathbf{Q}_{ti+1}^f = \mathbf{M}_{ti}^\top\mathbf{Q}_{ti}^a\mathbf{M}_{ti}.$$  

Proofs of all equations and a detailed treatment of the KF could be found in many statistical books, our notation conforms with e.g. that of Beezley [2009].

The KF is well known, its theoretical properties have been studied for a long time and under normality assumption and some other assumptions concerning independence of errors it has optimum properties. As such it is frequently used in many engineering applications. In Earth sciences, however, the length of the state vector is very high. For example in weather prediction, a 3D grid covering Europe in horizontal resolution of 10 km may have dimensions between 100 and 200 in each direction and e.g. 30 - 50 vertical levels. With 6 state variables this causes, that length of the model state vector $\mathbf{X}_{ti}$ is about $5 \times 10^6$. In such a case it is not possible to store the covariance matrices $\mathbf{Q}_{ti}^f$ and $\mathbf{Q}_{ti}^a$ in any computer memory and the formulas (6)–(8) start to be intractable in practice.
The Ensemble Kalman Filter

The basic idea behind the ensemble Kalman Filter is a low rank approximation of the covariance matrix $Q_f$. To define the EnKF we will have to restate some properties. At the time $t_i$ we will now work with a random sample $X_{f,t_i}^1, \ldots, X_{f,t_i}^N$, which in earth science is usually called "ensemble" and is often perceived as a set of possible scenarios, like possible evolutions of the atmospheric states. At the time $t_i$ we have the forecast ensemble, where each member of the ensemble is a column vector of size $n$ and $N$ is a number of ensemble members, $N \ll n$. Each ensemble member contains the whole model state. The analysis ensemble arises as result of application of formulas (6)–(8) on each of the members of the forecast ensemble, taking advantage of the low rank of the sample covariance matrix as a approximation to $Q_f$. We generate random perturbations of the input data

$$D_{t_i} = D_{t_i} + V_{t_i} \quad \forall j = 1, \ldots, N,$$

where $V_{t_i,j}$ are simulated from normal distribution independently of each other and of the forecast ensemble,

$$V_{t_i,j} \sim N(0, R_{t_i}) \quad \forall j = 1, \ldots, N.$$

These randomly perturbed data are then used to updating model state in analysis step.

Let $\bar{X}_{f,t_i}$ be the mean of the forecast ensemble

$$\bar{X}_{f,t_i} = \frac{1}{N} \sum_{j=1}^{N} X_{f,t_i}^j$$

and $C_{f,t_i}^f$ the sample covariance matrix

$$C_{f,t_i}^f = \frac{1}{N} \sum_{j=1}^{N} (X_{f,t_i}^j - \bar{X}_{f,t_i})(X_{f,t_i}^j - \bar{X}_{f,t_i})^\top.$$

The EnKF then replaces the true forecast mean by the mean of the forecast ensemble and the true covariance by the sample covariance matrix in Kalman update formulas (6)–(8). The formulas are applied on each ensemble member.

$$E_{t_i} = C_{f,t_i}^f H_{t_i}^\top (H_{t_i} C_{f,t_i}^f H_{t_i}^\top + R_{t_i})^{-1},$$

$$X_{t_i}^{a,j} = X_{f,t_i}^j + E_{t_i} (D_{t_i,j} - H_{t_i} X_{f,t_i}^j) \quad \text{(analysis step)},$$

$$X_{f,t_i+1}^j = M_{t_i} X_{t_i}^{a,j} + b_{t_i} \quad \text{(forecast step)}.$$

Matrix $E_{t_i}$ is called sample gain Kalman matrix and it is an approximation to Kalman gain matrix (6) used in the KF. It is important to realize, that while the forecast mean and covariance matrix in the KF were deterministic, the mean and sample covariance matrix used for EnKF are random quantities.

The crucial property of the EnKF is that we don’t need to store the matrices $C_{f,t_i}^f$ in memory of the computer. Let $u$ be any vector of dimension $n$, it easy to see that

$$C_{f,t_i}^f u = \left( \frac{1}{N} \sum_{j=1}^{N} (X_{f,t_i}^j - \bar{X}_{f,t_i})(X_{f,t_i}^j - \bar{X}_{f,t_i})^\top \right) u$$

$$= \frac{1}{N} \sum_{j=1}^{N} \left( (X_{f,t_i}^j - \bar{X}_{f,t_i})^\top u \right) (X_{f,t_i}^j - \bar{X}_{f,t_i}),$$

scalar product
so we simplify the computation of $C^f_t u$ to computation of $N$ scalar products of $n$-dimensional vectors, which is computationally feasible. Similarly we could simplify the computation of the sample Kalman gain matrix by using the equality

$$H_t^i C^f_t H_t^T = \frac{1}{N} \sum_{j=1}^{N} \left( H_t^i (X^f_{t,j} - \bar{X}^f_t) \right) \left( H_t^i (X^f_{t,j} - \bar{X}^f_t) \right)^T.$$ 

The rank of the sample covariance matrix $C^f_t$ is at most $N-1$ and thus the the perturbation of the forecast ensemble

$$X^a_{t,1} - X^f_{t,1}, \ldots, X^a_{t,N} - X^f_{t,N}$$

is contained in the space spanned by the columns of $C^f_t$ and any uncertainty outside of this subspace is simply ignored. It has been shown, for example by Anderson [2007], that this could cause the EnKF to diverge from the optimal solution. One of the methods, proposed to solve this problem, is called Localized Ensemble Filter and for more informations about it we recommend papers Anderson [2007] and Ott [2004].

It has been also shown that perturbations of the input data, in the presented form of the EnKF, brings artificially generated noise into the filter. There are versions of EnKF which avoid perturbations of the data. For example Evensen [2004] achieves this by adding another step into Kalman formula update

$$X^a_{t,i} = X^f_{t,i} + E_{t,i}(D_{t,i} - H_{t,i}X^f_{t,i}) \quad \text{(added step)},$$

$$X^a_{t+1,j} = M_{t,i}X^a_{t,i} + b_{t,i} \quad \text{(forecast step)},$$

where matrix $E_{t,i}$ is determined by solving the equation

$$C^a_{t,i} = (I - E_{t,i}H_{t,i})C^f_{t,i},$$

where $C^a_{t,i}$ is sample covariance matrix of the analysis ensemble $X^a_{t,1}, \ldots, X^a_{t,N}$. Detailed treatment of this situation is described in Beezley [2009].

Another large class of ensemble filters is formed by the so called square root filters which try to generate the perturbations of model states in some more efficient manner than purely random perturbations in the classical EnKF, for details see Tippett et al. [2003].

For a long time there was a lack of asymptotic properties of EnKF and many people were using it only by assuming that ensemble members are independent random vectors, which is not correct. This gap has been recently fulfilled mainly by papers Le Gland, Monbet and Tran [2009] and Mandel, Cobb and Beezley [2011]. The second paper is using a weak law of large numbers for exchangeable random variables (invariant under permutation) to prove that sample covariance matrix converges to true covariance matrix in probability as $N \to \infty$. Also $L^p$ bounds and convergence of ensemble members are proved. The first paper gives similar results, but the proofs are much more complicated. All convergences are meant under the assumption that $m, n$ are fixed, or bounded, and for $N \to \infty$.

The KF and EnKF on Hilbert spaces

The motivation for extending the EnKF to general Hilbert space (complete vector space with inner product) is to get more insight into the behaviour of the filter when the dimension $m$ of the input data grows. The asymptotic results, published in the cited papers, assumed that $m$ and $n$ were fixed or bounded. The question is, whether the convergence of EnKF to KF is not ruined by raising data dimension and whether is possible to find general convergence bound,
that could be applied on any Hilbert space.

First we need to define random variable on Hilbert spaces. Assume that $W$ is a infinite dimensional with inner product $\langle \cdot, \cdot \rangle$, we can define a random element on $W$ as a measurable function

$$X : (\Omega, \mathcal{S}, \mathbb{P}) \to (W, \mathcal{B}(W)),$$

where $(\Omega, \mathcal{S}, \mathbb{P})$ is a probability space and $\mathcal{B}(W)$ are Borel sets defined on $W$. It is also well known, see for example Bosq [2000], that mean value $E[X] \in W$ could be defined as a solution of equation

$$\langle u, E[X] \rangle = E[ \langle u, X \rangle ] \quad \forall u \in W.$$

This solution exists for all random elements from $L^1(\Omega, W)$ and it is unique. Similarly covariance operator could be defined.

In the ideal situation we would just apply Kalman update formula (6)–(8) to mean and covariance operator. However it brings many questions such as.

- is Bayes theorem still valid on infinite dimensional spaces? And how to define density on such spaces?
- When computing Kalman gain matrix (6), is expression $(H_t Q_t^{-1} H_t^T + R_t)^{-1}$ well defined? Doest it exist? If yes, is it bounded?
- When using the EnKF on $W$, how to define random perturbation?

Answers to these questions are still unknown and will be object of our next research.

References


