Convergence of the Ensemble Kalman Filter in Hilbert Space

Jan Mandel

Center for Computational Mathematics
Department of Mathematical and Statistical Sciences
University of Colorado Denver

Parts based on joint work with Loren Cobb and Jonathan Beezley
http://math.ucdenver.edu/~jmandel/papers/enkf-theory.pdf
http://arxiv.org/abs/0901.2951

Supported by NSF grant EGS-0835579 and Fondation STAE project ADTAO.

Toulouse, June 14, 2012
Data assimilation - update cycle

- Goal: Inject data into a running model

- Kalman filter used already in Apollo 11 navigation, now in GPS
Bayesian approach

- *discrete time filtering = sequential statistical estimation*
- Model state is a random variable $U$ with probability density $p(u)$
- Data + measurement error enter as data likelihood $p(d|u)$
- Probability densities updated by *Bayes theorem*: $p(u) \propto p(d|u)p_f(u)$
  $\propto$ means proportional

Bayesian update cycle

\begin{align*}
\text{time } k - 1 & \quad \text{advance model} & \text{time } k & \quad \text{Bayesian update} & \text{time } k \\
\text{analysis} & \quad U^{(k-1)} & \quad \text{forecast} & \quad p^{(k)}(u) & \quad \text{analysis} \\
p^{(k-1)}(u) & & p^{(k-1),f}(u) & & p^{(k)}(u)
\end{align*}

\[\text{data likelihood } p(d|u)\]
Kalman filter

• Model state $U$ is represented by its mean and covariance.
• Assume that the forecast is Gaussian, $U^f \sim N(m^f, Q^f)$, and
• the data likelihood is Gaussian,

$$p(d|u) \propto \exp[-(Hu - d)^T R^{-1} (Hu - d)/2].$$

• Then the analysis is also Gaussian, $U \sim N(m, Q)$, given by

$$m = m^f + K(d - Hm^f), \quad Q = (I - KH)Q^f$$
where $K = Q^f H^T (HQ^f H^T + R)^{-1}$ is the gain matrix
Kalman filter

- Hard to advance the covariance matrix when the model is nonlinear
- Modifications of the code needed – tangential and adjoint models
- Needs to maintain the covariance matrix - can’t do for large problems
- But the dimension of the state space is large - discretizations of PDEs
Ensemble Kalman filter (EnKF)

- represent model state by an ensemble, replace the forecast covariance by sample covariance computed from the forecast ensemble

\[
Q_N = \frac{1}{N} \sum_{i=1}^{N} (X_i^f - \overline{X}_N)(X_i^f - \overline{X}_N)^T, \quad \overline{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i^f
\]

- First idea (Evensen 1994): apply Kalman formula to the ensemble members:

\[
X_i = X_i^f + K_N(d - HX_i^f), \quad K_N = Q_N^f H^T(HQ_N^f H^T + R)^{-1}
\]

- But the analysis covariance was wrong – too small even if \( Q_N = Q \) were exact
- Fix: *perturb data by sampling from data error distribution*: replace \( d \) by \( D_i = d + E_i, \) i.i.d. \( E_i \sim N(0, R) \)
  (Burgers, Evensen, van Leeuwen, 1998)
The EnKF with data perturbation is correct if the covariance is

**Lemma.** Let $U^f \sim N(m^f, Q^f)$, $D = d + E$, $E \sim N(0, R)$ and

$$U = U^f + K(D - HU^f), \quad K = Q^f H^T (HQ^f H^T + R)^{-1}.$$ 

Then $U \sim N(m, Q)$ (i.e. $U$ has the correct analysis distribution from the Bayes theorem and the Kalman filter).


**Corollary.** If a forecast ensemble $U^f_i$ is a sample from the forecast distribution, and the exact covariance is used, then the analysis ensemble $U_i$ is a sample from the analysis distribution.
EnKF properties

- The model is needed only as a black box.
- The EnKF was derived for Gaussian distributions but the algorithm does not depend on this.
- So it is often used for **nonlinear models and non-Gaussian distributions** anyway.
- The EnKF was designed so that the ensemble is a sample from the correct analysis distribution if
  - the forecast ensemble is i.i.d. and Gaussian
  - the covariance matrix is exact
- But
  - the sample covariance matrix is a random element
  - the ensemble is not independent: sample covariance ties it together
  - the ensemble is not Gaussian: the update is a nonlinear mapping
Convergence of EnKF in the large ensemble limit


Here we simplify the approach from Mandel et al (2011) and show that its essence carries over to infinite dimension.
Convergence analysis overview

• For a large ensemble
  - the sample covariance converges to the exact covariance
  - the ensemble is asymptotically i.i.d.
  - the ensemble is asymptotically Gaussian
  - the ensemble members converge to the correct distribution

• How?
  - the ensemble is exchangeable: plays the role of independent
  - the ensemble is a-priori bounded in $L^p$: plays the role of Gaussian
  - use the weak law of large numbers for the sample covariance
  - the ensemble converges to a Gaussian i.i.d. ensemble obtained with
    the exact covariance matrix instead of the sample covariance

• Note: So far, we are still in a fixed, finite dimension. We’ll see later what survives in the infinitely dimensional case.
Notation: $|X|$ is the standard Euclidean norm in $\mathbb{R}^n$. If $X : \Omega \to \mathbb{R}^n$ is a random element, $\|X\|_p = E(|X|^p)^{1/p}$ is its stochastic $L^p$ norm, $1 \leq p < \infty$.

$X_j \xrightarrow{P} X$ (in probability) if $\forall \varepsilon > 0 : \Pr(|X_j - X| \geq \varepsilon) \to 0$.

Continuous mapping theorem: if $X_k \xrightarrow{P} X$ and $g$ is continuous, then $g(X_j) \xrightarrow{P} g(X)$.

$L^p$ convergence implies convergence in probability.
Tools

**Uniform integrability:** if \( X_j \overset{P}{\to} X \) and \( \{X_j\} \) is bounded in \( L^p \), then \( X_k \to X \) in \( L^q \) \( \forall \ 1 \leq q < p \).

Weak law of large numbers: if \( X_i \in L^2 \) are i.i.d., then \( \frac{1}{N} \sum_{i=1}^{N} X_i \to E(X_1) \) in \( L^2 \) as \( N \to \infty \).

Set of random elements \( \{X_1, \ldots, X_N\} \) is **exchangeable** if their joint distribution is invariant to a permutation of their indices. A set that is i.i.d. is exchangeable. Conversely, exchangeable random elements are identically distributed, but not necessarily independent.
Convergence of the EnKF to the Kalman filter

- Generate the initial ensemble $X_N^{(0)} = [X_{N1}^{(0)}, \ldots, X_{NN}^{(0)}]$ i.i.d. and Gaussian.
- Run EnKF, get ensembles $[X_N^{(k)}, \ldots, X_N^{(k)}]$, advance in time by linear models: $X_{Ni}^{k+1,f} = M_k X_{Ni}^k + f_k$.
- For theoretical purposes, define the reference ensembles $[U_{N1}^{(k)}, \ldots, U_{NN}^{(k)}]$ by $U_N^{(0)} = X_N^{(0)}$ and $U_N^{(k)}$ by the EnKF, except with the exact covariance of the filtering distribution instead of the sample covariance.
- We already know that $U_N^{(k)}$ is a sample (i.i.d.) from the Gaussian filtering distribution (as it would be delivered by the Kalman filter).

- We will show by induction that $X_{Ni}^{(k)} \to U_{Ni}^{(k)}$ in $L^p$ for all $1 \leq p < \infty$ as $N \to \infty$, for all analysis cycles $k$.
- In which sense? We need this for all $i = 1, \ldots, N$ but $U_{Ni}$ change and $N$ changes too.
Exchangeability of EnKF ensembles

**Lemma.** All ensembles \([X_{N1}^{(k)}, \ldots, X_{NN}^{(k)}]\) generated by the EnKF are exchangeable. The random elements \(\{\begin{bmatrix} X_{N1} \\ U_{N1} \end{bmatrix}, \ldots, \begin{bmatrix} X_{NN} \\ U_{NN} \end{bmatrix}\}\) are also exchangeable. (Recall that \(U_{Ni}\) are obtained using the exact covariance from the Kalman filter rather than the sample covariance.)

**Proof:** The initial ensemble \([X_{N1}^{(k)}, \ldots, X_{NN}^{(k)}] = [U_{N1}^{(k)}, \ldots, U_{NN}^{(k)}]\) is i.i.d., thus exchangeable, and the analysis step is permutation invariant.

**Corollary.** \([X_{N1} - U_{N1}, \ldots, X_{NN} - U_{NN}]\) are identically distributed, so we only need to consider the convergence \(\|X_{N1} - U_{N1}\|_p \to 0\).
Estimates for the covariance

Write the \( N \)-th sample covariance of \([X_1, X_2, \ldots]\), as

\[
C_N(X) = \frac{1}{N} \sum_{i=1}^{N} X_i X_i^T - \left( \frac{1}{N} \sum_{i=1}^{N} X_i \right) \left( \frac{1}{N} \sum_{i=1}^{N} X_i \right)^T
\]

**Bound** If \( X_i \in L^{2p} \) are identically distributed, then

\[
\|C_N(X)\|_p \leq \|X_1 X_1^T\|_p + \|X_1\|_p^2 \leq \|X_1\|_{2p}^2 + \|X_1\|_p^2
\]

**Continuity of the sample covariance:** If \([X_i, U_i] \in L^4\) are identically distributed, then

\[
\|C_N(X) - C_N(U)\|_2 \leq \sqrt{8} \|X_1 - U_1\|_4 \sqrt{\|X_1\|_4^2 + \|U_1\|_4^2}
\]
Estimates for the covariance

$L^2$ law of large numbers for the sample covariance: If $U_i \in L^4$, $i = 1, 2, \ldots$ are i.i.d., then

$$\|C_N(X) - C_N(U)\|_2 \to 0, N \to \infty.$$ 

Proof. $X_iX_i^T \in L^2$ are i.i.d. and the weak law of large numbers applies.

Corollary If $[X_{Ni}, U_i] \in L^4$ are identically distributed, $U_i$ are independent, and $\|X_{N1} - U_1\|_4 \to 0$ as $N \to \infty$, then $C_N(X) \to \text{Cov}(U_1)$ in $L^2$, thus also

$$C_N(X) \overset{p}{\to} \text{Cov}(U_1).$$
A-priori $L^p$ bounds on EnKF ensembles

Lemma. All EnKF ensembles $[X_N^{(k)}, \ldots, X_N^{(k)}]$ are bounded in $L^p$, $\|X_{N1}^{(k)}\|_p \leq C_{k,p}$ with constants independent of the ensemble size $N$, for all $1 \leq p < \infty$.

Proof: By induction over step index $k$.
- Recall that
  
  \[X_N^{(k)} = X_N^{(k),f} + K_N^{(k)}(D_N^{(k)} - H^{(k)}X_N^{(k),f})\]
  \[K_N^{(k)} = Q_N^{(k),f}H^{(k)}(H^{(k)}Q_N^{(k),f}H^{(k)} + R^{(k)})^{-1}\]

- Bound the sample covariance $\|Q_N^{(k),f}\|_p$ from the $L^{2p}$ bound on $X_N^{(k),f}$
- Bound the matrix norm $|(H^{(k)}Q_N^{(k),f}H^{(k)} + R^{(k)})^{-1}| \leq \text{const}$ from $R^{(k)} > 0$ (positive definite).
Convergence of the EnKF is now simple:

Recall: $X_{Ni}^{(k)} = \text{EnKF ensemble}$, $U_i^{(k)} = \text{reference i.i.d. ensemble from exact covariances}$. By induction over step index $k$:

• $X_{N1}^{(k),f} \rightarrow U_1^{(k),f}$ in $L^p$ as $N \rightarrow \infty$, $\forall 1 \leq p < \infty$

• Sample covariance $Q_N^{(k),f} = C_N \left( X_N^{(k),f} \right) \xrightarrow{P} Q(k),f = \text{Cov} \left( U_1^{(k),f} \right)$

• Continuous mapping theorem $\implies$ Kalman gain $Q_N^{(k)} \xrightarrow{P} K(k)$ and $X_{N1}^{(k)} \xrightarrow{P} U_1^{(k)}$.

• Linearly advancing in time: $X_{N1}^{(k+1),f} \xrightarrow{P} U_1^{(k+1),f}$

• A-priori $L^p$ bound and uniform integrability $\implies X_{N1}^{(k+1),f} \rightarrow U_1^{(k+1),f}$ in $L^p \forall 1 \leq p < \infty$
Towards high-dimension asymptotics

• We have proved convergence in large sample limit for arbitrary but fixed state space dimension. This says nothing about the speed of convergence as the state dimension increases.
• Curse of dimensionality: slow convergence in high dimension. Why and when is it happening?
• With arbitrary probability distributions, sure. But probability distributions in practice are not arbitrary: the state is a discretization of a random function.
• Random functions are random elements with values in infinitely dimensional spaces. State dimension is just from the mesh size.
• The smoother the functions, the faster the EnKF convergence.
• As a step towards convergence uniform in state space dimension, look at the infinitely dimensional case first - just like in numerical analysis where we study the PDE first a PDE and only then its numerical approximation
Bayes theorem in an infinitely dimensional Hilbert space

Model state is a probability measure $\mu$ on a \textbf{separable} Hilbert space $V$ with inner product $\langle \cdot, \cdot \rangle$.

Avoid densities: Bayes theorem $p(u) \propto p(d\mid u) p_f(u)$ becomes

$$
\mu(A) = \frac{\int_A p(d\mid u) du(\mu_f)}{\int_V p(d\mid u) du(\mu_f)} \quad \forall \mu_f\text{-measurable } A
$$

Need $\int_H p(d\mid u) du(\mu_f) > 0$ for the Bayes theorem to work. Data likelihood $p(d\mid u) > 0$ for all $u$ from some open set in $V$ is enough.

Gaussian data likelihood:

$$
p(d\mid u) = \exp\left(-\frac{1}{2} \|R^{-1/2}(Hu-d)\|^2\right) \text{ if } Hu-d \in \text{Im}(R^{1/2}), 0 \text{ otherwise.}
$$

In particular, $R = I$ works just fine. But $R$ cannot be covariance of a probability measure if the data space is infinitely dimensional.
Probability on an infinitely dimensional Hilbert space

The mean of a random element $X$ is the weak integral:

$$E(X) \in V, \quad \langle v, E(X) \rangle = E(\langle v, X \rangle) \quad \forall v \in V.$$

Tensor product of vectors $x, y \in V$ was $xy^\top$, now it is the bounded linear operator

$$x \otimes y \in [V], \quad (x \otimes y) v = x \langle y, v \rangle \quad \forall v \in V$$

Covariance of a random element $X$ becomes

$$\text{Cov}(X) = E((X - E(X)) \otimes (X - E(X)))$$
$$= E(X \otimes X) - E(X) \otimes E(X)$$

Covariance of a random element on $V$ must be a compact operator of trace class: its eigenvalues $\lambda_n$ must satisfy $\sum_{n=1}^{\infty} \lambda_n < \infty$. 
Curse of dimensionality? Not for probability measures!

Constant covariance eigenvalues $\lambda_n = 1$ and the inverse law $\lambda_n = 1/n$ are not probability measures in the limit because $\sum_{n=1}^{\infty} \lambda_n = \infty$.

Inverse square law $\lambda_n = 1/n^2$ gives a probability measure because $\sum_{n=1}^{\infty} \lambda_n < \infty$.

$m=25$ uniformly sampled data points from 1D state, $N=10$ ensemble members. From Beezley (2009), Fig. 4.7.
The $L^2$ and the weak laws of large numbers

The definition of $L^p(\Omega, V) = \{u : E(|u|^p < \infty)\}$ and $\|u\|_p$ carry over.

The $L^2$ law of large numbers for sample mean $E_N = \frac{1}{N} \sum_{i=1}^{N} X_i$ carries over: Let $X_k \in L^2(\Omega, V)$ be i.i.d., then

$$\|E_N - E(X_1)\|_2 \leq \frac{1}{\sqrt{N}} \|X_1 - E(X_1)\|_2 \leq \frac{2}{\sqrt{N}} \|X_1\|_2,$$

and consequently $E_N \overset{P}{\rightarrow} E(X_1)$.

But $L^p$ laws of large numbers do not generally hold on Banach spaces (in fact define Rademacher type $p$ spaces), and the space $[V]$ of all bounded linear operators on a Hilbert space is not a Hilbert space but a quite general Banach space.
Sample covariance

The definition carries over from finite dimension:

\[ C_N ([X_1, \ldots, X_N]) = \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i - \left( \frac{1}{N} \sum_{i=1}^{N} X_i \right) \otimes \left( \frac{1}{N} \sum_{i=1}^{N} X_i \right) \]

But we do not have the \( L^2 (\Omega, [V]) \) law of large numbers for the sample covariance. Instead use \( L^2 (\Omega, V_{HS}) \) with \( V_{HS} \) the space of all Hilbert-Schmidt operators on \( V \) and the norm

\[ |A|_{HS}^2 = \sum_{n=1}^{\infty} \langle Ae_n, Ae_n \rangle \]

where \( \{e_n\} \) is any complete orthonormal sequence in \( V \). Now the \( L^2 \) law of large numbers applies in the Hilbert space \( V_{HS} \), and \( |A|_{[V]} \leq |A|_{HS} \), so we recover

\[ C_N ([X_1, \ldots, X_N]) \to \text{Cov} X_1 \text{ in } L^2 (\Omega, [V]). \]
Convergence of EnKF in a Hilbert space

A-priori $L^p$ bounds on EnKF ensembles: We need only

$$\left| \left( H^{(k)} Q_N^{(k)}, f H^{(k)\ast} + R^{(k)} \right)^{-1} \right| < \text{const}$$

then the bound caries over. This is true when the data error covariance has spectrum bounded below, $R^{(k)} \geq cI > 0$ - such as $R^{(k)} = I$. This was good for the Bayes theorem already.

With these few changes the whole proof carries over, avoiding entry-by-entry arguments. But not all is good - the formulation of the EnKF algorithm now involves data perturbation by white noise. And white noise is not a random element with values in the Hilbert space, at least not in the usual sense.
EnKF in a Hilbert space with white noise data perturbation

A weak random element (Balakrishnan 1976) has finitely additive – not necessarily σ-additive distribution, which allows identity covariance (white noise). If the covariance of a gaussian weak random element is of trace class, the distribution is in fact σ-additive and the weak random element is a random element in the usual sense.

So let $D = d + E$, $E$ is weak random element $\sim N(0, R)$, $U_f \sim N(m, Q_f)$ and

$$U = U_f + K(D - HU), \quad \text{with} \quad K = QH^*(H^*QH^* + R)^{-1}$$

$\implies$ the analysis $U$ is also a weak random variable with the analysis distribution, $N(m, Q)$, $Q = (I - KH)Q$ trace class

$\implies$ the distribution of the analysis $U$ is a gaussian probability measure.
Conclusion

- With a bit of care, a simplified proof in finite dimension carries over to infinitely dimensional separable Hilbert space.
- White noise data error is good. White noise state space is bad.
- Computational experiments confirm that EnKF converges uniformly for high-dimensional distributions that approximate a Gaussian measure on Hilbert space. (J. Beezley, Ph.D. thesis, 2009). EnKF for distributions with slowly decaying eigenvalues converges very slowly and requires huge ensembles.
- The same is seen for particle filters... in spite of the fact that particle filters for problems without a structure fail in high dimension.
- Square root EnKF has no data perturbation - but the ensemble may not be exchangeable.
- Next: uniform convergence for high-dimensional problems that approximate an infinitely dimensional one ... just like numerical PDEs?
References


