Convergence of ensemble Kalman filters in the large ensemble limit and infinite dimension

Jan Mandel

Department of Mathematical and Statistical Sciences

University of Colorado Denver

Based on joint work with Loren Cobb, Evan Kwiatkowski, and Kody Law

Supported by NSF grant DMS-1216481

Colorado School of Mines, Golden, CO, April 24, 2015

Data assimilation - Analysis cycle

- Goal: Inject data into a running model

	analysis step		advance in time		analysis step		
forecast	$\longrightarrow$	analysis	$\longrightarrow$	forecast	$\longrightarrow$	analysis	...
data	$\nearrow$			data	$\nearrow$		

- Kalman filter used already in Apollo 11 navigation, now in GPS, computer vision, weather forecasting, remote sensing,...

From least squares to Bayes theorem

Inverse problem $Hu \approx d$ with background knowledge $u \approx u^f$

$$|u - u^f|^2_{Q^{-1}} + |d - Hu|^2_{R^{-1}} \to \min_u$$

u^f=forecast, what we think the state should be, d=data,
H=observation operator, Hu=what the data should be given state u.

$$\underbrace{e^{-\frac{1}{2}\left(|u-u^f|^2_{Q^{-1}}+|d-Hu|^2_{R^{-1}}\right)}}_{\substack{\propto\, p^a(u)=p(u|d) \\ \text{analysis (posterior) density}}} = \underbrace{e^{-\frac{1}{2}|u-u^f|^2_{Q^{-1}}}}_{\substack{\propto\, p(u) \\ \text{forecast (prior) density}}} \underbrace{e^{-\frac{1}{2}|d-Hu|^2_{R^{-1}}}}_{\substack{\propto\, p(d|u) \\ \text{data likelihood}}} \to \max_u$$

This is Bayes theorem for probability densities. $\propto$ means proportional.

The analysis density $p^a(u) \propto e^{-\frac{1}{2}|u-u^a|^2_{Q^{a-1}}}$ mean and covariance are

$u^a = u^f + K(d - Hu^f), \quad Q^a = (I - KH)Q$,
$K = \mathcal{K}(Q) = QH^\mathsf{T}(HQH^\mathsf{T} + R)^{-1}$ is the Kalman gain

Kalman filter (KF)

• Add advancing in time - model is operator $u \mapsto Mu + b$:
mean and covariance advanced as Mu and MQM^{T}
• Hard to advance the covariance matrix when the model is nonlinear.
• Need tangent and adjoint model operators.
• Can't maintain the covariance matrix for large problems, such as from discretizations of PDEs.

Ensemble forecasting to the rescue

• Ensemble weather forecast: Independent randomized simulations. If it rains in 30% of them, then say that "the chance of rain is 30%".

• Ensemble Kalman filter (EnKF, Evensen 1994): replace the covariance in KF by the sample covariance of the ensemble, then apply the KF formula for the mean to each ensemble member independently.

• But the analysis covariance was wrong – too small even if the covariance were exact.

• Fix: *randomize also the data by sampling from the data error distribution*. (Burgers, Evensen, van Leeuwen, 1998).

• Then all is good and we should get the right ensembles...
at least in the Gaussian case, and up to the approximation by the sample covariance.

The EnKF with data randomization is statistically exact if the covariance is exact

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

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

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

Proof. Computation in Burgers, Evensen, van Leeuwen, 1998.

Corollary. If the forecast ensemble $[U_1^f, \ldots, U_N^f]$ is a sample from Gaussian forecast distribution, and the exact covariance is used, then the analysis ensemble $[U_1^a, \ldots, U_N^a]$ is a sample from the analysis distribution.

EnKF properties

- The model is needed only as a **black box**.
- The ensemble members interact through ensemble mean and covariance only
- 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.
- Folklore: "high-dimensional problems require large ensembles" a.k.a. "**curse of dimensionality**". Not necessarily so.
- Curse of dimensionality: slow convergence in high dimension. With arbitrary probability distributions, sure. But probability distributions in practice are not arbitrary: the state is a discretization of a random function.The smoother the functions, the faster the EnKF convergence.

Convergence of EnKF in the large ensemble limit

• Laws of large numbers to guarantee that the EnKF gives correct results for large ensembles, in the Gaussian case: Le Gland et al. (2011), Mandel et al (2011).

• In general, the EnKF converges to a mean-field limit (Le Gland et al. 2011).

- mean-field approximation = the effect of all other particles on any one particle is replaced by a single averaged effect
- mean field limit = large number of particles, the influence of each becomes negligible.

For convergence in high state space dimension, we look at the infinite dimension first - just like in numerical analysis where we study first the PDE and only then its numerical approximation

Tools: Random elements on a separable Hilbert space

The mean of a random element X on Hilbert space V:

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

Similarly for random bounded operator A on V,

$$E(A) \in V, \quad \langle v, E(A)u\rangle = E(\langle v, Au\rangle) \quad \forall u, v \in V$$

Tensor product of vectors xy^{T}, x, $y \in V$, becomes

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

Covariance of a random element X becomes

$$\begin{aligned} C = \mathsf{Cov}(X) &= E((X - E(X)) \otimes (X - E(X))) \\ &= E(X \otimes X) - E(X) \otimes E(X) \end{aligned}$$

Covariance must be a compact operator with finite trace,
$\mathsf{Tr}\, C = \sum_{n=1}^{\infty} \lambda_n < \infty$ (λ_n eigenvalues).

Tools: Karhunen-Loève expansion and random functions

Suppose $X \in L^2(\Omega, V) = \left\{X : E\left(|X|^2\right) < \infty\right\}$. Then $C = \operatorname{Cov}(X)$compact and self-adjoint, has complete orthonormal system $\{e_n\}$ of eigenvectors, $Ce_n = \lambda_n e_n$, eigenvalues $\lambda_n \geq 0$, and

$$X = E(X) + \textstyle\sum_{n=1}^{\infty} \lambda_n^{1/2} \xi_n e_n, \quad E(\xi_n) = 0.$$

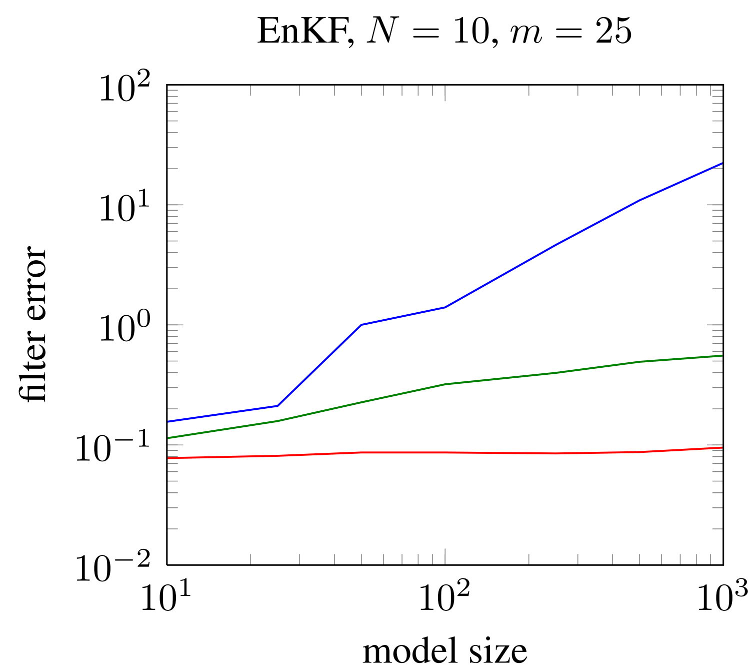
is convergent a.s. in V, and ξ_m are uncorrelated random variables $\langle \xi_m, \xi_n \rangle_{L^2(\Omega)} = E(\xi_m, \xi_n) = \delta_{mn}$. Examples:

$$X = \left(\lambda_1^{1/2}\xi_1, \lambda_2^{1/2}\xi_2, \ldots\right) \text{ in } \ell^2,$$

$$X(x) = \sum_{n=1}^{\infty} \lambda_n^{1/2} \xi_n \cos(nx) \text{ in } L^2(0, \pi)$$

with $\xi_k \sim N(0, 1)$ independent, $\lambda_n = n^{-\alpha}$, $\alpha > 1$.

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 the thesis Beezley (2009), Fig. 4.7.

Bayes theorem in infinite dimension

- Forecast and analysis are probability distributions, densities p^f, p^a
- Bayes theorem: $p^a(u) = \frac{1}{C} p(d|u)\, p^f(u)$, $C = \int\limits_U p(d|u)\, p^f du$
- Write in terms of measures μ^a, μ^f with densities p^a, p^f

$$\int\limits_A p^a(u)\, du = \frac{1}{C} \int\limits_A p(d|u)\, p^f(u)\, du,\ C = \int\limits_U p(d|u)\, d\mu^f(u)$$

$$\mu^a(A) = \frac{1}{C} \int\limits_A p(d|u)\, d\mu^f(u),\ \forall A\ \mu^f\text{-measurable}$$

$\Leftrightarrow$ Radon-Nikodym derivative: $p(d|u) = C \dfrac{d\mu^a}{d\mu^f}$

- **But how do we know that** $C = \int\limits_U p(d|u)\, d\mu^f(u) > 0$?

Gaussian case

- data likelihood: $d = Hu + \varepsilon$, $\varepsilon \sim N(0, R)$,

$p(d|u) = \text{const}\, e^{-\frac{1}{2}|Hu-d|^2_{R^{-1}}}$

- $\mu^f(u)$ is Gaussian measure on U, data $d \in V$

- state space U and data space V are separable Hilbert spaces

- $p(d|\cdot) \in L^1\left(U, \mu^f\right) \Rightarrow \mu^a$ is absolutely continuous w.r.t. μ^f

- Feldman-Hájek theorem$\Rightarrow$ μ^a and μ^f equivalent

- Difficulties when the data are infinitely dimensional...

Infinite-dimensional data, Gaussian measure error bad

- Example: $\mu^f = N(0, Q)$, $H = I \Leftrightarrow$ all state observed, $d = 0$, $R = Q$
- $p(d|u) = \text{const}\, e^{-\frac{1}{2}|u|^2_{R^{-1}}} = \text{const}\, e^{-\frac{1}{2}\left\langle R^{-1/2}u, R^{-1/2}u\right\rangle}$
- $p(d|u) > 0$ if $u \in R^{1/2}(U) = \text{dom}\left(R^{-1/2}\right)$
- $p(d|u) = e^{-\infty} = 0$ if $u \notin R^{1/2}(U) = Q^{1/2}(U)$
- $Q^{1/2}(U)$ is the Cameron-Martin space of the measure $N(0, Q)$
- But $\mu = N(0, Q) \Rightarrow \mu\left(Q^{1/2}(U)\right) = 0$. Thus, $\int_U p(d|u)\, d\mu^f(u) = 0$
- Also not possible from the Feldman-Hájek theorem,
  μ^a and μ^f are not equivalent.

Infinite-dimensional data, white noise error good

- All is good when data is finite-dimensional and R not singular

- More generally, when data covariance R is bounded below:

$$\langle u, Ru \rangle \geq \alpha |u|^2 \ \forall u,\ \alpha > 0 \Rightarrow |u|^2_{R^{-1}} < \infty \ \forall u \ \Rightarrow e^{-|Hu-d|^2_{R^{-1}}} > 0 \ \forall u$$

$$\Rightarrow \int_U p(d|u)\, d\mu^f(u) = \text{const} \int_U e^{-|Hu-d|^2_{R^{-1}}} d\mu^f(u) > 0$$

- But if V is not finite dimensional, then **such data error distribution $N(0, R)$ is not a probability measure on** V. Covariance of probability a measure has finite trace $\sum_{n=1}^{\infty} \lambda_n < \infty$, but the spectrum $\sigma(R) \geq \alpha > 0$.

- Yet Bayes theorem $\mu^a(A) = \frac{1}{C} \int_A p(d|u)\, d\mu^f(u)$ is fine.

Randomized data ensemble Kalman filter

Initial ensemble: $\left[Y_1^0, Y_2^0, \ldots, Y_N^0\right]$ i.i.d., Gaussian bakground

Advance time by the model: $X_k^m = F\left(Y_k^{m-1}\right)$

Analysis step: nonlinear map

$\left[X_1^m, X_2^m, \ldots, X_N^m\right] \mapsto \left[Y_1^m, Y_2^m, \ldots, Y_N^m\right]$

by solving the inverse problem $HY \approx d^m$

$$Y_k^m = X_k^m + \mathcal{K}(Q_N^m)(D_k^m - HX_k^m)$$

$$Q_N^m = C_N\left[X_1^m, \ldots, X_N^m\right],\ D_k^m \sim N\left(d^m, R\right)$$

Randomized data EnKF and white data noise

- $Y_k^m = X_k^m + \mathcal{K}(Q_N^m)(D_k^m - HX_k^m)$, $Q_N^m = C_N\left[X_1^m, \ldots, X_N^m\right]$, $D_k^m \sim N\left(d^m, R\right)$

- Data space V infinitely dimensional, $R \geq \alpha I$, $\alpha > 0$
  $\Rightarrow N\left(d^m, R\right)$ is only finitely additive on the algebra of cylindrical sets, not a σ additive measure on the σ-algebra of Borel sets of V
  D_k^m are only **weak** random variables (linear functional is a r.v)
  But $\mathrm{Tr}\left(Q^{1/2}\right) < \infty \Rightarrow Q^{1/2}\cdot$bounded operator$\cdot D_k^m$
  has σ-additive distribution and is a random element
  $\Rightarrow \mathcal{K}(Q_N^m)D_k^m = QH^{\mathsf{T}}(HQH^{\mathsf{T}} + R)^{-1}D_k^m$
  has σ-additive distribution and is a random element

- Back to standard Kolmogorov probability theory

Tools: L^p law of large numbers on a Hilbert space

Define $L^p(\Omega, V) = \{X : E(|X|^p) < \infty\}$ equipped with the norm

$$\|X\|_p = E(|X|^p)^{1/p}$$

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

$$\|E_N - E(X_1)\|_p \leq \frac{C_p}{\sqrt{N}} \|X_1\|_p .$$

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**. Worse, any Banach space can be embedded in $[V]$ for some Hilbert space V. So, if something does not hold on Banach spaces, it won't be true for bounded operators on Hilbert spaces either.

Tools: L^p law of large numbers for the sample covariance

Let $X_i \in L^p(\Omega, V)$ be i.i.d. and

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

But $[V]$ is not a Hilbert space.
Use $L^2(\Omega, V_{HS})$ with V_{HS} the space of Hilbert-Schmidt operators,

$$|A|_{HS}^2 = \sum_{n=1}^{\infty} \langle Ae_n, Ae_n \rangle < \infty, \quad \langle A, B \rangle_{HS} = \sum_{n=1}^{\infty} \langle Ae_n, Be_n \rangle,$$

where $\{e_n\}$ is any complete orthonormal sequence in V. In finite dimension, the Hilbert-Schmidt norm becomes the Frobenius norm

$$|A|_{HS}^2 = \sqrt{\sum_{i,j} \left|a_{ij}\right|^2}.$$

Tools: L^p law of large numbers for the sample covariance

Let $X_i \in L^p(\Omega, V)$ be i.i.d. and V a separable Hilbert space.

Then, from the L^p law of large numbers in the Hilbert space V_{HS} of Hilbert-Schmidt operators on V,

$$\|C_N([X_1, \ldots, X_N]) - \mathsf{Cov}(X_1)\|_p \leq \||C_N([X_1, \ldots, X_N]) - \mathsf{Cov}(X_1)|_{\mathsf{HS}}\|_p$$
$$\leq \frac{\mathsf{const}(p)}{\sqrt{N}} \|X_1\|_{2p}^2 \ .$$

where the constant depends on p only.

Note that since V is separable, V_{HS} is also separable.

Convergence of the EnKF to mean-field limit

• Legland et al. (2011): analysis step as nonlinear transformation of probability measures.

• Nonlinear tranformation of **ensemble as vector of exchangeable random variables** $[X_1, X_2, \ldots, X_N] \mapsto [Y_1, Y_2, \ldots, Y_N]$

• L^p continuity of the model. Drop the superscript m

• Mean field filter:
$Y_k^{\text{Mean field}} = X_k^{\text{Mean field}} + \mathcal{K}(Q)(D_k - HX_k^{\text{Mean field}})$, $Q = \text{Cov}\,(X_1)$

• Randomized EnKF:
$$Y_k^{\text{Randomized}} = X_k^{\text{Randomized}} + \mathcal{K}(Q_N)(D_k - HX_k^{\text{Randomized}}),$$
$$Q_N = C_N\left[X_1^{\text{Randomized}}, \ldots, X_N^{\text{Randomized}}\right], \text{ same } D_k$$

Convergence of the EnKF to mean-field limit (cont)

- Subtract, continuity of Kalman gain:
$\|\mathcal{K}(Q) - \mathcal{K}(Q_N)\|_p \leq \text{const}\, \|Q - Q_N\|_{2p}$

- L^p law of large numbers for $C_N\left[X_1^{\text{Mean field}}, \ldots, X_N^{\text{Mean field}}\right]$.

- Apriori bound $\left\|X_k^m\right\|_p \leq \text{const}\,(m)$ for all m from
$\left|(HQH^* + R)^{-1}\right| \leq \frac{1}{\alpha}$ by $R \geq \alpha I$

- induction over m: $X_1^{m,\text{Randomized}} \rightarrow X_1^{m,\text{Mean field}}$ in all L^p,
$1 \leq p < \infty$

Conclusion

• With a bit of care, a convergence proof in finite dimension caries over to infinitely dimensional separable Hilbert space.

• White noise data error is good, stabilizes the method

• But white noise distribution is not a standard σ-additive probability measure.

• Convergence in non–linear non-Gaussian case to mean-field limit

• Square root EnKF has no randomization, convergence in Gaussian case Kwiatkowski and M. 2015

• 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 of the covariance converges very slowly and requires large ensembles.

References

[1] A. V. Balakrishnan. *Applied functional analysis*. Springer-Verlag, New York, 1976.

[2] Jonathan D. Beezley. *High-Dimensional Data Assimilation and Morphing Ensemble Kalman Filters with Applications in Wildfire Modeling*. PhD thesis, University of Colorado Denver, 2009.

[3] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.

[4] Giuseppe Da Prato. *An introduction to infinite-dimensional analysis*. Springer-Verlag, Berlin, 2006.

[5] Geir Evensen. Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 (C5)(10):143–162, 1994.

[6] Evan Kwiatkowski and Jan Mandel. Convergence of the square root ensemble Kalman filter in the large ensemble limit. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1–17, 2015.

[7] Kody J. H. Law, Hamidou Tembine, and Raul Tempone. Deterministic methods for filtering, part I: Mean-field ensemble Kalman filtering. arXiv:1409.0628, 2014.

[8] F. Le Gland, V. Monbet, and V.-D. Tran. Large sample asymptotics for the ensemble Kalman filter. In Dan Crisan and Boris Rozovskiĭ, editors, *The Oxford Handbook of Nonlinear Filtering*, pages 598–631. Oxford University Press, 2011.

[9] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Vol. 23. Springer-Verlag, Berlin, 1991.

[10] Jan Mandel, Loren Cobb, and Jonathan D. Beezley. On the convergence of the ensemble Kalman filter. *Applications of Mathematics*, 56:533–541, 2011.

[11] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.