On the Convergence of the Ensemble Kalman Filter

Jan Mandel[a,*], Loren Cobb[a], Jonathan Beezley[a]

[a]*Center for Computational Mathematics and Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO 80217-3364*

Abstract

Convergence of the ensemble Kalman filter in the limit for large ensembles to the Kalman filter is proved. In each step of the filter, convergence of the ensemble sample covariance follows from a weak law of large numbers for exchangeable random variables, Slutsky's theorem gives weak convergence of ensemble members, and L^p bounds on the ensemble then give L^p convergence.

Key words: Exchangeable random variables, Monte-Carlo methods, data assimilation, theoretical analysis, asymptotics, EnKF, filtering

1. Introduction

Data assimilation, a topic of importance in many disciplines, uses statistical estimation to update the state of a running model based on new data. One of the most succesful recent data assimilation methods is the ensemble Kalman filter (EnKF). EnKF is a Monte-Carlo approximation of the Kalman filter (KF), with the covariance in the KF replaced by the sample covariance computed from an ensemble of realizations. Because the EnKF does not need to maintain the state covariance matrix, it is suitable for high-dimensional problems.

A large body of literature on the EnKF and variants exists, but rigorous probabilistic analysis is lacking. It is commonly assumed that the ensemble is a sample (that is, i.i.d.) and it is normally distributed. Although the resulting analyses played an important role in the development of EnKF, both assumptions are false. The ensemble covariance is computed from all ensemble members together, thus introducing dependence, and the EnKF formula is a nonlinear function of the ensemble, thus destroying the normality of the ensemble distribution.

The present analysis does not employ these two assumptions. The ensemble members are shown to be exchangeable random variables bounded in L^p, which provides properties that replace independence and normality. An argument using uniform integrability and Slutsky's theorem is then possible. The result is valid for the EnKF version of Burgers, van Leeuven, and Evensen

*Corresponding author
Email addresses: jan.mandel@gmail.com (Jan Mandel)

(Burgers et al., 1998) in the case of constant state space dimension, a linear model, normal data likelihood and initial state distributions, and ensemble size going to infinity. This EnKF version involves randomization of data. Efficient variants of EnKF without randomization exist (Anderson, 1999; Tippett et al., 2003), but they are not the subject of this paper.

The analysis in Burgers et al. (1998) consists of the comparison of the covariance of the analysis ensemble and the covariance of the filtering distribution under the assumption that the ensemble covariance converges in the limit for large ensembles. Furrer and Bengtsson (2007) note that if the ensemble sample covariance is a consistent estimator, then Slutsky's theorem yields the convergence in probability of the gain matrix. When this article was being completed, we became aware of a presentation by Le Gland which announces related results but does not seem to take advantage of exchangeability.

2. Preliminaries

The Euclidean norm of column vectors in $\mathbb{R}^m$, $m \geq 1$, and the induced matrix norm are denoted by $\|\cdot\|$, and $^{\mathrm{T}}$ is the transpose. The (stochastic) L^p norm of random element is $\|X\|_p = (E(\|X\|^p))^{1/p}$. The j-th entry of a vector X is $[X]_j$ and the i,j entry of a matrix $Y \in \mathbb{R}^{m\times n}$ is $[Y]_{ij}$ .Weak convergence (convergence in distribution) is denoted by $\Rightarrow$; weak convergence to a constant is the same as convergence in probability. *All convergence is for $N \to \infty$.* We denote by $X_N = [X_{Ni}]_{i=1}^N = [X_{N1}, \ldots, X_{NN}]$, with various superscripts and for various $m \geq 1$, an ensemble of N random elements, called members, with values in $\mathbb{R}^m$. Thus, an ensemble is a random $m \times N$ matrix with the ensemble members as columns. Given two ensembles X_N and Y_N, the stacked ensemble $[X_N; Y_N]$ is defined as the block random matrix

$$[X_N; Y_N] = \begin{bmatrix} X_N \\ Y_N \end{bmatrix} = \left[\begin{bmatrix} X_{N1} \\ Y_{N1} \end{bmatrix}, \ldots, \begin{bmatrix} X_{NN} \\ Y_{NN} \end{bmatrix} \right] = [X_{Ni}; Y_{Ni}]_{i=1}^N.$$

If all the members of X_N are identically distributed, we write $E(X_{N1})$ and $\mathrm{Cov}(X_{N1})$ for their common mean vector and covariance matrix. The ensemble sample mean and ensemble sample covariance matrix are the random elements $\overline{X}_N = \frac{1}{N}\sum_{i=1}^N X_{Ni}$ and $C(X_N) = \overline{X_N X_N^{\mathrm{T}}} - \overline{X}_N \overline{X}_N^{\mathrm{T}}$.

We will work with ensembles such that the joint distribution of the ensemble X_N is invariant under a permutation of the ensemble members. Such ensemble is called exchangeable. An ensemble X_N is exchangeable if and only if $\Pr(X_N \in B) = \Pr(X_N \Pi \in B)$ for every Borel set $B \subset \mathbb{R}^{m\times N}$ and every permutation matrix $\Pi \in \mathbb{R}^{N\times N}$. The covariance between any two members of an exchangeable ensemble is the same, $\mathrm{Cov}(X_{Ni}, X_{Nj}) = \mathrm{Cov}(X_{N1}, X_{N2})$, $i \neq j$.

Lemma 1. *Suppose X_N and D_N are exchangeable, the random elements X_N and D_N are independent, and $Y_{Ni} = F(X_N, X_{Ni}, D_{Ni})$, $i = 1, \ldots, N$, where F is measurable and permutation invariant in the first argument, i.e.*

$F(X_N\Pi, X_{Ni}, D_{Ni}) = F(X_N, X_{Ni}, D_{Ni})$ *for any permutation matrix* Π*. Then* Y_N *is exchangeable.*

PROOF. Write $Y_N = \mathbf{F}(X_N, D_N)$, where

$$\mathbf{F}(X_N, D_N) = [F(X_N, X_{N1}, D_{N1}), \ldots, F^{(k)}(X_N, X_{NN}, D_{NN})].$$

Let Π be a permutation matrix. Then $Y_N\Pi = \mathbf{F}(X_N\Pi, D_N\Pi)$. Because X_N is exchangeable, the distributions of X_N and $X_N\Pi$ are identical. Similarly, the distributions of D_N and $D_N\Pi$ are identical. Since X_N and D_N are independent, the joint distributions of (X_N, D_N) and $(X_N\Pi, D_N\Pi)$ are identical. Thus, for any Borel set $B \subset \mathbb{R}^{n\times N}$, $\Pr(Y_N\Pi \in B) = E(1_B(Y_N\Pi)) = E(1_B(\mathbf{F}(X_N\Pi, D_N\Pi))) = E(1_B(\mathbf{F}(X_N, D_N))) = \Pr(X_N \in B)$. $\square$

We now prove a weak law of large numbers for exchangeable ensembles.

Lemma 2. *If for all* N*,* X_N*,* U_N *are ensembles of* $\mathbb{R}^1$ *valued random variables,* $[X_N; U_N]$ *is exchangeable,* $\operatorname{Cov}(U_{Ni}, U_{Nj}) = 0$ *for all* $i \neq j$*,* $U_{N1} \in L^2$ *is the same for all* N*, and* $X_{N1} \to U_{N1}$ *in* L^2*, then* $\overline{X}_N \Rightarrow E(U_{N1})$*.*

PROOF. Since X_N is exchangeable, $\operatorname{Cov}(X_{Ni}, X_{Nj}) = \operatorname{Cov}(X_{N1}, X_{N2})$ for all $i, j = 1, \ldots, N$, $i \neq j$. Since $X_N - U_N$ is exchangeable, also $X_{N2} - U_{N2} \to 0$ in L^2. Then, using the identity $\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y)$ and Cauchy inequality for the L^2 inner product $E(XY)$, we have $|\operatorname{Cov}(X_{N1}, X_{N2}) - \operatorname{Cov}(U_{N1}, U_{N2})| \leq 2\|X_{N1}\|_2\|X_{N2} - U_{N2}\|_2 + 2\|U_{N2}\|_2\|X_{N1} - U_{N1}\|_2$, so $\operatorname{Cov}(X_{N1}, X_{N2}) \Rightarrow 0$. By the same argument, $\operatorname{Var}(X_{N1}) \Rightarrow \operatorname{Var}(U_{N1}) < +\infty$. Now $E(\overline{X}_N) = E(X_{N1}) \Rightarrow E(U_{N1})$ from $X_{N1} - U_{N1} \to 0$ in L^2, and $\operatorname{Var}(\overline{X}_N) = \frac{1}{N^2}\sum_{i=1}^N \operatorname{Var}(X_{Ni}) + \sum_{i,j=1, j\neq j}^N \operatorname{Cov}(X_{Ni}, X_{Nj}) = \frac{1}{N}\operatorname{Var}(X_{N1}) + (1 - \frac{1}{N})\operatorname{Cov}(X_{N1}, X_{N2}) \to 0$, and the conclusion follows from Chebyshev inequality. $\square$

The convergence of the ensemble sample covariance for nearly i.i.d. exchangeable ensembles follows.

Lemma 3. *If for all* N*,* X_N*,* U_N *are ensembles of* $\mathbb{R}^n$ *valued random elements,* $[X_N; U_N]$ *is exchangeable,* U_N *are i.i.d.,* $U_{N1} \in L^4$ *is the same for all* N*, and* $X_{N1} \to U_{N1}$ *in* L^4*, then* $\overline{X}_N \Rightarrow E(U_{N1})$ *and* $C(X_N) \Rightarrow \operatorname{Cov}(U_{N1})$*.*

PROOF. From Lemma 2, it follows that $[\overline{X}_N]_j \Rightarrow [E(U_{N1})]_j$ for each entry $j = 1, \ldots, n$, so $\overline{X}_N \Rightarrow E(U_{N1})$. Let $Y_{Ni} = X_{Ni}X_{Ni}^{\mathrm{T}}$, so that $C(X_N) = \overline{Y}_N - \overline{X}_N\overline{X}_N^{\mathrm{T}}$. Each entry of $[Y_{Ni}]_{j\ell} = [X_{Ni}]_j[X_{Ni}]_\ell$ satisfies the assumptions of Lemma 2, so $[Y_{Ni}]_{j\ell} \Rightarrow E([U_{N1}U_{N1}^{\mathrm{T}}]_{j\ell})$. Convergence of the entries $[\overline{X}_N\overline{X}_N^{\mathrm{T}}]_{j\ell} = [\overline{X}_N]_j[\overline{X}_N]_\ell$ to $E([U_{N1}]_{j\ell})E([U_{N1}^{\mathrm{T}}]_{j\ell})$ follows from the already proved convergence of $\overline{X}_N$ and Slutsky's theorem (Chow and Teicher, 1997, p. 254). Applying Slutsky's theorem again, we get $C(X_N) \Rightarrow \operatorname{Cov}(U_{N1})$. $\square$

3. Formulation of the EnKF

Consider an initial state given as the random variable $U^{(0)}$. In step k, the state $U^{(k-1)}$ is advanced in time by applying the model $M^{(k)}$ to obtain $U^{(k),f} = M^{(k)}(U^{(k-1)})$, called the prior or the forecast, with probability density function (pdf) $p_{U^{(k),f}}$. The data in step k are given as measurements $d^{(k)}$ with a known error distribution, and expressed as the data likelihood $p(d^{(k)}|u)$. The new state $U^{(k)}$ conditional on the data, called the posterior or the analysis, then has the density $p_{U^{(k)}}$ given by the Bayes theorem, $p_{U^{(k)}}(u) \propto p(d^{(k)}|u)p_{U^{(k),f}}(u)$, where $\propto$ means proportional. This is the discrete time filtering problem. The distribution of $U^{(k)}$ is called the filtering distribution.

Assume $U^{(0)} \sim N(u^{(0)}, Q^{(0)})$, the model is linear, $M^{(k)} : u \mapsto A^{(k)}u + b^{(k)}$, and the data likelihood is normal, $d^{(k)} \sim N(H^{(k)}u^{(k),f}, R^{(k)})$ given $u^{(k),f}$, where $H^{(k)}$ is the given observation matrix and $R^{(k)}$ is the given data error covariance, and the data error is independent of the model state. Then the filtering distribution is normal, $U^{(k)} \sim N(u^{(k)}, Q^{(k)})$, and it satisfies the KF recursions (Anderson and Moore, 1979)

$$u^{(k),f} = E(U^{(k),f}) = A^{(k)}u^{(k)} + b^{(k)}, \quad Q^{(k),f} = \operatorname{Cov} U^{(k),f} = A^{(k)^{\mathrm{T}}} Q^{(k)} A^{(k)},$$
$$u^{(k)} = u^{(k),f} + L^{(k)}(d^{(k)} - H^{(k)}u^{(k),f}), \quad Q^{(k)} = (I - L^{(k)}H^{(k)})Q^{(k),f},$$

where the Kalman gain matrix $L^{(k)}$ is given by

$$L^{(k)} = Q^{(k),f} H^{(k)\mathrm{T}} (H^{(k)} Q^{(k),f} H^{(k)\mathrm{T}} + R^{(k)})^{-1}. \tag{1}$$

The EnKF is essentially based on the following observation. Let $U_i^{(0)} \sim N(u^{(0)}, Q^{(0)})$ and $D_i^{(k)} \sim N(d^{(k)}, R^{(k)})$ be independent for all $k, i \geq 1$. Given N, choose the initial ensemble and the perturbed data as the the first N terms of the respective sequence, $U_{Ni}^{(0)} = U_i^{(0)}$, $i = 1, \ldots, N$, $D_{Ni}^{(k)} = D_i^{(k)}$, $i = 1, \ldots, N$, $k = 1, 2, \ldots$ Define the ensembles $U_N^{(k)}$ by applying the KF formulas to each ensemble member separately using the corresponding member of perturbed data,

$$U_{Ni}^{(k),f} = M^{(k)}(U_{Ni}^{(k-1)}), \quad i = 1, \ldots, N, \tag{2}$$
$$U_N^{(k)} = U_N^{(k),f} + L^{(k)}(D_N^{(k)} - H^{(k)}U_N^{(k),f}). \tag{3}$$

The next lemma shows that $U_N^{(k)}$ is a sample from the filtering distribution.

Lemma 4. *For all $k = 1, 2, \ldots$, U_N^k is i.i.d. and $U_{N1}^{(k)} \sim N(u^{(k)}, Q^{(k)})$.*

Proof. The statement is true for $k = 0$ by definition of $U_N^{(0)}$. Assume that it is true for $k-1$ in place of k. The ensemble $U_N^{(k)}$ is i.i.d. and normally distributed because it is an image under a linear map of the normally distributed i.i.d. ensemble with members $[U_{Ni}^{(k-1)}, D_{Ni}^{(k)}]$, $i = 1, \ldots, N$. Further, $D_N^{(k)}$ and $U_{Ni}^{(k),f}$ are independent, so from Burgers et al. (1998, eq. (15) and (16)), $U_{N1}^{(k)}$ has the correct mean and covariance, which determines the normal distribution of $U_{N1}^{(k)}$ uniquely. □

The EnKF is now obtained by replacing the exact covariance $L^{(k)}$ by the ensemble sample covariance. The ensembles produced by EnKF are $X_N^{(0)} = U_N^{(0)}$ and

$$X_{Ni}^{(k),f} = \ M^{(k)}(X_{Ni}^{(k-1)}), \quad i = 1, \ldots, N. \tag{4}$$

$$X_N^{(k)} = X_N^{(k),f} + K_N^{(k)}(D_N^{(k)} - H^{(k)} X_N^{(k),f}), \tag{5}$$

where $K_N^{(k)}$ is the ensemble sample gain matrix,

$$K_N^{(k)} = Q_N^{(k)} H^{(k)T} (H^{(k)} Q_N^{(k)} H^{(k)T} + R^{(k)})^{-1}, \quad Q_N^{(k)} = C(X_N^{(k),f}). \tag{6}$$

4. Convergence analysis

Lemma 5. *There exist constants $c(k,p)$ for all k and all $p < \infty$ such that $\|X_{N1}^{(k)}\|_p \le c(k,p)$ and $\|K_N^{(k)}\|_p \le c(k,p)$ for all N.*

Proof. For $k = 0$, each $X_{Ni}^{(k)}$ is normal. Assume $\|X_{N1}^{(k-1)}\|_p \le c(k-1,p)$ for all N. Then

$$\|X_{N1}^{(k),f}\|_p = \|A^{(k)} X_{N1}^{(k-1)} + b^{(k)}\|_p \le \|A^{(k)}\| \|X_{N1}^{(k-1)}\|_p + \|b^{(k)}\| \le \mathrm{const}(k,p).$$

By Jensen's inequality, for any X_N, $\|\frac{1}{N}\sum_{i=1}^N X_{N_i}\|_p \le \frac{1}{N}\sum_{i=1}^N \|X_{N_i}\|_p$. This gives $\|\overline{X}_N^{(k),f}\|_p \le \mathrm{const}(k,p)$ and $\|Q_N^{(k)}\|_p \le \|X_{N1}^{(k),f} X_{N1}^{(k),f\mathrm{T}}\|_p + \|X_{N1}^{(k),f}\|_p^2 \le \|X_{N1}^{(k),f}\|_{2p}^2 + \|X_{N1}^{(k),f}\|_p^2 \le \mathrm{const}(k,p)$, since from Cauchy inequality,

$$\|WZ\|_p \le E\left(\|W\|^p \|Z\|^p\right)^{\frac{1}{p}} \le E(\|W\|^{2p})^{\frac{1}{2p}} E(\|Z\|^{2p})^{\frac{1}{2p}} = \|W\|_{2p} \|Z\|_{2p}, \tag{7}$$

for any compatible random matrices W and Z.

Since $H^{(k)} Q_N^{(k)} H^{(k)\mathrm{T}}$ is symmetric positive semidefinite and $R^{(k)}$ is symmetric positive definite, it holds that$\|(H^{(k)} Q_N^{(k)} H^{(k)\mathrm{T}} + R^{(k)})^{-1}\| \le \|(R^{(k)})^{-1}\| \le \mathrm{const}(k)$, which, together with the bound on $\|Q_N^{(k)}\|_p$, gives $\|K_N^{(k)}\|_p \le \|Q_N^{(k)}\|_p \,\mathrm{const}(k) \le \mathrm{const}(k,p)$. Finally, we obtain the desired bound

$$\begin{aligned}\|X_{N1}^{(k)}\|_p &\le \|X_{N1}^{(k),f}\|_p + \|K_N^{(k)} D_{N1}^{(k)}\|_p + \|K_N^{(k)} H^{(k)} X_{N1}^{(k),f}\|_p \\ &\le \mathrm{const}(k,p)(\|X_{N1}^{(k),f}\|_p + \|K_N^{(k)}\|_p + \|K_N^{(k)}\|_{2p} \|X_{N1}^{(k),f}\|_{2p}) \le c(k,p),\end{aligned}$$

using again (7). □

Theorem 1. *For all k, $[X_N; U_N]$ is exchangeable and $X_{N1}^{(k)} \to U_{N1}^{(k)}$ in L^p for all $p < +\infty$.*

Proof. The ensembles $U_N^{(k)}$ are obtained by linear mapping of the i.i.d. initial ensemble $U_N^{(0)}$, so they are i.i.d. Since $X_{Ni}^{(0)} = U_{Ni}^{(0)}$, $[X_N^{(0)}; U_N^{(0)}]$ is exchangeable, and $X_{N1} = U_{N1}$. Suppose the statement holds for $k-1$ in place of k. The ensemble members are given by a recursion of the form $[X_{Ni}^{(k)}; U_{Ni}^{(k)}] = F^{(k)}(C(X_N^{(k-1)}), [X_{Ni}^{(k-1)}; U_{Ni}^{(k-1)}], D_{Ni}^{(k)})$. The ensemble sample covariance matrix C is permutation invariant, so $[X_N^{(k)}; U_N^{(k)}]$ is exchangeable by Lemma 1. Subtracting (5) and (3) gives $X_N^{(k),f} - U_N^{(k),f} = A^{(k)}(X_N^{(k-1)} - U_N^{(k-1)})$, and $X_N^{(k),f}$ and $U_N^{(k),f}$ satisfy the assumption of Lemma 3. Thus, $C(X_N^{(k),f}) \Rightarrow \operatorname{Cov} U_{N1}^{(k),f}$, $K_N^{(k)} \Rightarrow L^{(k)}$ by the mapping theorem (Billingsley, 1995, p. 334), and $X_{N1}^{(k)} \Rightarrow U_{N1}^{(k)}$ by Slutsky's theorem. Since for all $p < +\infty$, the sequence $\{X_{N1}^{(k)}\}_{N=1}^{\infty}$ is bounded in L^p by Lemma 5 and $U_{N1}^{(k)} \in L^p$, it follows that $X_{N1}^{(k)} \to U_{N1}^{(k)}$ in L^p for all $p < +\infty$ from uniform integrability (Billingsley, 1995, p. 338). $\square$

Using Lemma 3 and uniform integrability again, it follows that the ensemble mean and covariance are consistent estimators of the filtering mean and covariance.

Corollary 1. *$\overline{X}_N^{(k)} \to u^{(k)}$ and $C(X_N^{(k)}) \to Q^{(k)}$ in L^p for all $p < +\infty$.*

5. Acknowledgements

This work was partially supported by the National Science Foundation under grants CNS-0719641 and ATM-0835579. We would like to thank Professor Richard Bradley for an example that led us away from mixing coefficients, and Dr. Thomas Bengtsson for bringing Furrer and Bengtsson (2007) to our attention.

References

Anderson, B. D. O., Moore, J. B., 1979. Optimal filtering. Prentice-Hall, Englewood Cliffs, N.J.

Anderson, J. L., 1999. An ensemble adjustment Kalman filter for data assimilation. Monthly Weather Review 129, 2884–2903.

Billingsley, P., 1995. Probability and measure, 3rd Edition. John Wiley & Sons Inc., New York.

Burgers, G., van Leeuwen, P. J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Monthly Weather Review 126, 1719–1724.

Chow, Y. S., Teicher, H., 1997. Probability theory, 3rd Edition. Springer-Verlag, New York.

Furrer, R., Bengtsson, T., 2007. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivariate Anal. 98 (2), 227–255.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., Whitaker, J. S., 2003. Ensemble square root filters. Monthly Weather Review 131, 1485–1490.