Advances and Applications in Statistics
Volume 32, Number 2, 2013, Pages 151-162
Published Online: February 2013
Available online at http://pphmj.com/journals/adas.htm
Published by Pushpa Publishing House, Allahabad, INDIA

EXPLORING EnKF CONVERGENCE FOR LINEAR DYNAMICAL SYSTEM

Nina Fitriyati[1,4]**, Sutawanir Darwis**[1]**, Agus Yodi Gunawan**[2] **and Asep Kurnia Permadi**[3]

[1]Statistics Research Division
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung
Indonesia
e-mail: nina_fitriyati@students.itb.ac.id

[2]Industrial and Financial Mathematics Research Division
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung
Indonesia

[3]Reservoir Engineering Research Division
Faculty of Petroleum Engineering
Institut Teknologi Bandung
Indonesia

[4]Mathematics Department
Faculty of Sciences and Technology
UIN Syarif Hidayatullah Jakarta
Indonesia

2010 Mathematics Subject Classification: 62M20.

Keywords and phrases: ensemble Kalman filter, convergence in probability, control chart, Chebyshev's inequality.

Communicated by K. K. Azad

Received November 23, 2012

Abstract

The discussion of convergence is important to justify the validity of the ensemble Kalman filter (EnKF) method. In linear system, it is well known theoretically that for the very big number of the ensemble sizes, the state in EnKF hardly differs with the state in Kalman filter. In this paper, we explore numerically such convergence by means of the statistical quality control chart methodology. If the random samples of the average of the forecast step and the update step in EnKF are viewed as production outputs of a process with specific characteristics, namely, the forecast step and the update step in Kalman filter, we can observe their behaviors around these characteristics using $\bar{x}$ control chart. We propose the control limits as a function of the ensemble size which is calculated based on the Chebyshev's inequality. Simulation study shows that, as the ensemble size increases, the band of control limits will be tighter and the average of state in the forecast step and the update step of the EnKF will be closed to the same Kalman filter quantities. This simulation confirms theoretically the convergence of EnKF to Kalman filter.

Introduction

The EnKF is a popular sequentially data assimilation technique, since it was first introduced by Evensen [3]. EnKF is a Monte Carlo implementation of Kalman filter for nonlinear dynamical system where initial state is generated using sample, called as *ensemble*, and the error covariance matrix is approximated using empirical covariance matrix. To justify the validity of EnKF, it is important to investigate the convergence in probability of the EnKF. Asymptotic convergence of the EnKF for a linear dynamic system has been discussed by researchers. Butala et al. [1] represented that, for a large ensemble, EnKF estimate converge in probability to an optimal linear estimation minimum mean square error (LMMSE) is given by Kalman filter estimate. The similar results are shown by Tan [11]. If the element of the ensembles are viewed as a population of particle filter with additive Gaussian white noise and Gaussian initial condition, the probability distribution of random samples of the average of the forecast step and the update step will coincide with the Gaussian distributions associated with the Kalman predictor and with the Kalman filter [6]. And if the states are viewed as an exchangeable random variables bounded in L^p, $p \in [1, \infty)$, Mandel et al. [8] proved that EnKF converges to the Kalman filter by means of weak convergence and L^p convergence.

Let $\overline{X}_k^{f,N}$ and $\overline{X}_k^{a,N}$ be the averages of the state in the forecast step and the update step in EnKF for *k*th cycle with ensemble size *N*, respectively. Similar to this, $\widehat{P}_k^{f,N}$ and $\widehat{P}_k^{a,N}$ denote the covariance matrices in the forecast and the update step in EnKF. Now, let the corresponding quantities for these variables in Kalman filter be given by X_k^f, P_k^f, X_k^a and P_k^a:

$$\overline{X}_k^{f,N} \xrightarrow{p} X_k^f \quad \text{and} \quad \widehat{P}_k^{f,N} \xrightarrow{p} P_k^f, \tag{1}$$

$$\overline{X}_k^{a,N} \xrightarrow{p} X_k^a \quad \text{and} \quad \widehat{P}_k^{a,N} \xrightarrow{p} P_k^a. \tag{2}$$

The convergence of (1) and (2) in EnKF by means of numerical errors update has been discussed in Li and Xiu [7], for a nonlinear operator state, the local truncation error was bounded by statistical errors and numerical error of the algorithm for the forecast model.

Many researchers developed methodologies to understand the convergence concepts. Bryce et al. [2] advised more work on computer methods to explain the convergence concepts in statistics. A visual-minded and graphical simulation-based approach was proposed by Lafaye De Micheaux and Liquet [5] to describe the concept of convergence in probability, convergence almost surely, convergence in law, and convergence in *r*th mean and to explain the convergence in law and Marasinghe et al. [9] combined the multiple simulations and high resolution dynamic graphics to describe the convergence in law. In this paper, we aim to improve the understanding of convergence in probability of the EnKF. We will explore this convergence by means of the statistical quality control chart methodology. If the random samples of the average of the forecast step and the update step in EnKF are viewed as production outputs of a process with specific characteristics, namely, the forecast step and the update step in Kalman filter, we can analyze their behavior around these characteristics using $\bar{x}$ control chart. We propose the control limits as a function of the ensemble size which is calculated based on the Chebyshev's inequality.

Methodology

In this paper, we consider a linear dynamical system where the state-space are represented by

$$\begin{cases} \text{State}: X_k = FX_{k-1}^a + \eta_{k-1},\ E(\eta_k) = 0,\ E(\eta_k\eta_k^T) = Q_k \\ \text{Measurement}: Y_k = HX_k + \varepsilon_k,\ E(\varepsilon_k) = 0,\ E(\varepsilon_k\varepsilon_k^T) = R_k, \end{cases} \tag{3}$$

where F and H are state and measurement operators, respectively, η_k and ε_k are model error and measurement error, respectively. Assume that η_k and ε_k are independent at all lags. Let $D_0 = \varnothing$ and $D_k = \{Y_1^0, Y_2^0, ..., Y_k^0\}$ be the set of all measurements that are available until time k and $X_0^a = E(X_0 | D_0)$ is given initial state where error covariance matrix is $P_0^a = E((X_0^a - X_0)(X_0^a - X_0)^T)$. The basic equations in the Kalman filter are

1. In the forecast step

$$\text{State: } X_k^f = FX_{k-1}^a$$

$$\text{Covariance matrix: } P_k^f = FP_{k-1}^aF^T + Q_k. \tag{4}$$

2. In the update step

$$\text{State: } X_k^a = E(X_k | D_k) = X_k^f + K_k(Y_k - HX_k^f)$$

$$\text{Covariance matrix: } P_k^a = (I - K_kH)P_k^f, \tag{5}$$

where $K_k = P_k^f H^T(HP_k^f H^T + R_k)^{-1}$ is the Kalman gain matrix.

The EnKF is a Monte-Carlo implementation of Kalman filter where initial ensemble is generated by adding a sequence of independent identically distributed (iid) of error models to X_0^a. Let N be the ensemble size. The initial state is generated by $X_{0,i}^a = X_0^a + \eta_{0,i}$, $i = 1, 2, ..., N$, where $\eta_0 \sim N(0, P_0^a)$. The ensembles in the forecast step are $X_{k,i}^f = FX_{k-1,i}^a$, the empirical mean vector is $\overline{X}_k^{f,N} = F\overline{X}_{k-1}^{a,N}$ and the empirical covariance matrix is

$$\widehat{P}_k^{f,N} = \frac{1}{N-1}\sum_{i=1}^{N}(X_{k,i}^{f,N} - \overline{X}_k^{f,N})(X_{k,i}^{f,N} - \overline{X}_k^{f,N})^T.$$

The ensembles in the update step are $X_{k,i}^a = X_{k,i}^f + \widehat{K}_k(Y_k - HX_{k,i}^f)$ and the empirical mean vector is $\overline{X}_k^{a,N} = \overline{X}_k^{f,N} + \widehat{K}_k(Y_k - H\overline{X}_k^{f,N})$, where $\widehat{K}_k = \widehat{P}_k^{f,N}H^T(H\widehat{P}_k^{f,N}H^T + R_k)^{-1}$ is the Kalman gain matrix and $Y_k = Y_k^0 + \varepsilon_k$, $\varepsilon_k \sim N(0, R_k)$.

Without loss of generality, we consider visualizing the convergence of the scalar system of F and H in the EnKF. Let M be the replications of sampling of size N and $\overline{X}_{k,1}^{f,N}\overline{X}_{k,2}^{f,N}, ..., \overline{X}_{k,M}^{f,N}$ be the iid random vectors of the empirical means of the forecast step in EnKF for kth cycle. From equation (1), we have $\lim_{N\to\infty} P(|\overline{X}_{k,j}^{f,N} - X_k^f| > \varepsilon) = 0$, for $\forall\varepsilon > 0$, $k > 0$ and $j = 1, 2, ..., M$ or equivalent to $p_N^f = P(|\overline{X}_{k,j}^{f,N} - X_k^f| > \varepsilon) \underset{N\to\infty}{\to} 0.$ We use the Frequentist approach (Fisher [4]) to approximate the probability p_N^f by the proportion

$$\widehat{p}_N^f = \#\{|\overline{x}_{k,j}^{f,N} - x_k^f| > \varepsilon\}/M,$$

where $\#\{|\ \bar{x}_{k,j}^{f,N} - x_k^f\ | > \varepsilon\}$ are the numbers of successful events that $|\ \bar{x}_{k,j}^{f,N} - x_k^f\ | > \varepsilon$. Choose $\varepsilon = c\sqrt{P_k^f}/\sqrt{N} > 0$, where c is a multiplier of standard deviation of Chebyshev's inequality. The probability bound of these iid random vectors follows the Chebyshev's inequality:

$$P(|\ \bar{X}_{k,j}^{f,N} - X_k^f\ | > c\sqrt{P_k^f/N}) < \frac{1}{c^2},\ c \in \mathbb{R}. \tag{6}$$

Using the same quantity above, in the update step, we have the probability bound follows Chebyshev's inequality:

$$P(|\ \bar{X}_{k,j}^{a,N} - X_k^a\ | > c\sqrt{P_k^a/N}) < \frac{1}{c^2},\ c \in \mathbb{R} \tag{7}$$

and the proportion $\hat{p}_N^a = \#\{|\ \bar{x}_{k,j}^{a,N} - x_k^a\ | > \varepsilon\}/M$. Using equation (7), we can determine the ensemble size N as a function of ε, i.e., $N = c^2 P_k^a/\varepsilon^2$.

Here, we consider the visualization of the convergence of the EnKF using statistical quality control charts methodology. The $\bar{x}$ control chart is widely used to monitor mean of variables that consists of three limits: upper line (UL), centre line (CL), and lower line (LL) where $LL < CL < UL$ [10]. We use this chart to visualize the iid random vectors of the average of the forecast step and the update step in the EnKF methodology and investigate behaviors of these iid random vectors around the corresponding quantity in Kalman filter when the ensemble size increases. Based on equations (6) and (7), for the control limits in the forecast step, we choose $UL = X_k^f + c\sqrt{P_k^f/N}$, $CL = X_k^f$, and $LL = X_k^f - c\sqrt{P_k^f/N}$, and for the update step, we choose $UL = X_k^a + c\sqrt{P_k^a/N}$, $CL = X_k^a$, and $LL = X_k^a - c\sqrt{P_k^a/N}$. From these control limits, we observe that the band between UL and LL will be tighter when the ensemble size increases. Therefore, this visualization improves the understanding of the convergence of EnKF.

Results and Discussions

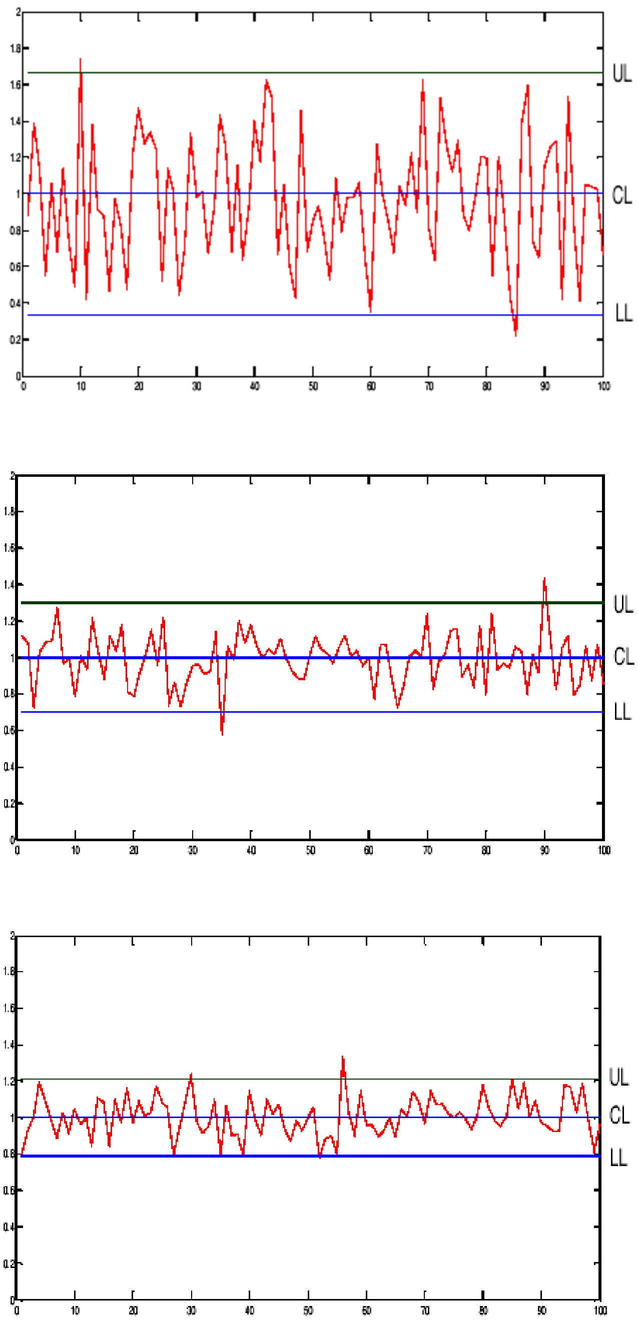
The simulation is purposed to improve the understanding of the convergence in probability of EnKF for the forecast step and the update step. Consider state space in the random walks model as:

$$\begin{cases} X_k = X_{k-1} + \eta_{k-1}, \ \eta_k \sim N(0, 1), \\ Y_k = X_k + \varepsilon_k, \ \varepsilon_k \sim N(0, 2). \end{cases}$$

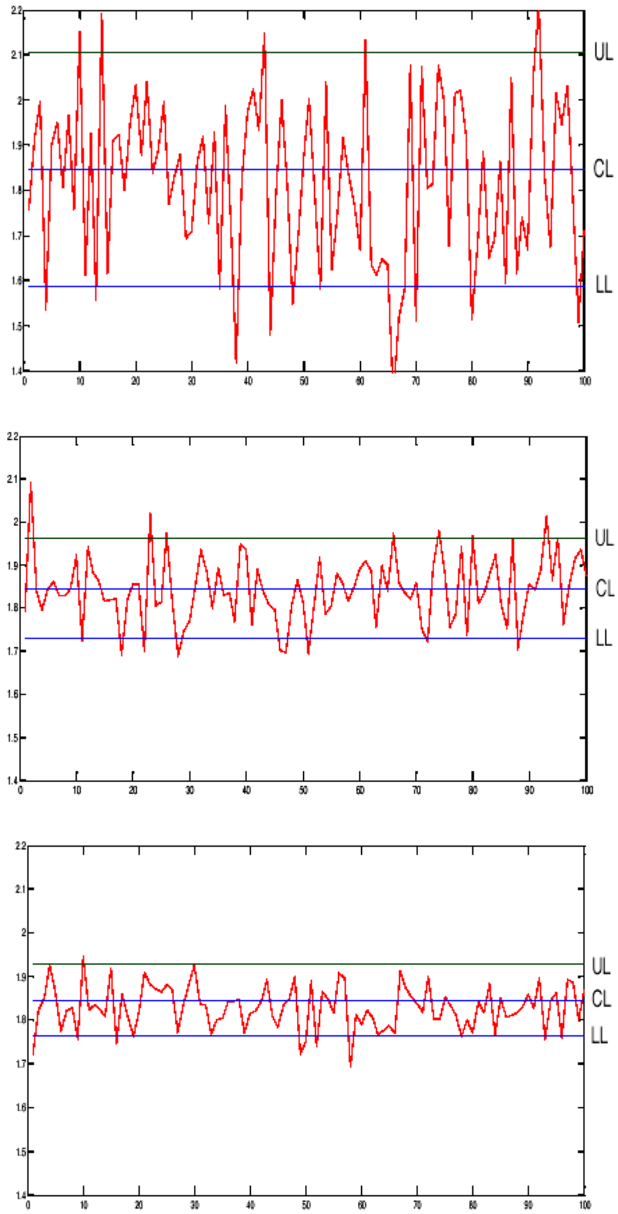
For illustration, we take $X_0^a = 1$, $P_0^a = 10$ and $Y_1^0 = 2$. Using Kalman filter algorithm, equations (4) to (5), we have $P_1^f = 11$, $K_1 = \frac{11}{13}$, $X_1^a = \frac{24}{13}$, $P_1^a = \frac{22}{13}$. Consider N initial ensembles in the EnKF algorithm which are generated from $X_{0,i}^a \sim N(X_0^a, P_0^a)$, $i = 1, 2, ..., N$. For fix replications M and multiplier of Chebyshev's inequality c, we vary the size of ensemble N. Here, we take $M = 100$, $c = 2$ and ensemble sizes 100, 500 and 1000. The $\bar{x}$ chart is shown in Figures 1A (forecast step) and 1B (update step). From these figures, we observe that, from $M = 100$ replications, less than 25% of the forecast step and the update step layout of the control limits. Increasing the ensemble size N, we find that the band of control limits becomes narrow. From these figures, we can also observe that the value of $\varepsilon = c\sqrt{P_k^a}/\sqrt{N}$ will be smaller when the ensemble size increases. The various numbers of the ensemble sizes as a function of ε are shown in Table 1. This table shows that as $\varepsilon \to 0$, the $N \to \infty$.

Table 1. The ensemble sizes for the various value of ε

ε	0.1	0.075	0.05	0.01	0.001
N	677	1.203	2.708	67.692	6.769.231

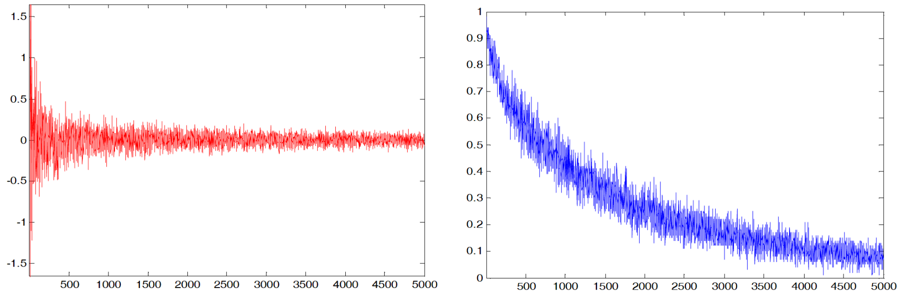


A



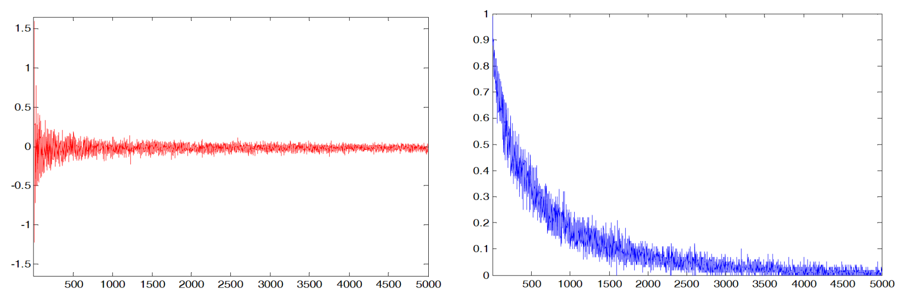
B

Figure 1. The $\bar{x}$ chart for A. The average of the forecast step and B. The average of the update step in the 1st cycle in EnKF with three ensemble sizes, 100 (upper), 500 (middle), and 1000 (lower).



A B

Figure 2. A. The pattern of $Y_k^f = \overline{X}_k^{f,N} - X_k^f$ for the ensemble size $N = 1$ to 5000, B. The proportion of p_N^f in $[-\varepsilon, \varepsilon] = [-0.08, 0.08]$ toward to 0 as the ensemble size increases.



A B

Figure 3. A. The pattern of $Y_k^a = \overline{X}_k^{a,N} - X_k^a$ for the ensemble size $N = 1$ to 5000, B. The proportion of p_N^a in $[-\varepsilon, \varepsilon] = [-0.08, 0.08]$ toward to 0 as the ensemble size increases.

Let $Y_k^{f,N} = \overline{X}_k^{f,N} - X_k^f$ and $Y_k^{a,N} = \overline{X}_k^{a,N} - X_k^a$ be the difference between EnKF and Kalman filter for the forecast step and the update step, respectively. Figures 2A and 3A show the pattern of Y_k^f and Y_k^a, respectively. As the ensemble size increases, the difference between EnKF and Kalman filter for average of the forecast step and the update step will close to zero. Therefore, the average of forecast step and the update step of the EnKF will be closed to the same Kalman filter quantity. The evolution of $\hat{p}_n^f$ and $\hat{p}_n^a$ are shown in Figures 2B and 3B. From these figures, we observe that $\hat{p}_n^f$ and $\hat{p}_n^a$ toward zero when the ensemble size increases. These figures confirm the convergence in probability of the forecast step and the update step in EnKF as shown in equations (1) and (2), theoretically.

Conclusions

Investigation of convergence is important to justify the validity of EnKF method. In this paper, we visualize the convergence in probability of the EnKF by means of the statistical quality control chart methodology. We propose the $\bar{x}$ control limits as a function of the ensemble size which is calculated based on the Chebyshev's inequality. Simulation study of random walk case showed that the band of control limits becomes narrow when the ensemble size increases and the difference between EnKF and Kalman filter for average of the forecast step and the update step will close to zero when the ensemble size is very big. This visualization improves understanding of EnKF convergence.

References

[1] M. D. Butala, J. Yun, Y. Chen, R. A. Frazin and F. Kamalabadi, Asymptotic convergence of the ensemble Kalman filter, 15th IEEE International Conference on Image Processing, 2008.

[2] G. R. Bryce, R. Gould, W. Notz and R. Peck, Curriculum guidelines for bachelor of science degrees in statistical science, Amer. Statist. 55 (2001), 7-13.

[3] G. Evensen, Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research 99(C5) (1994), 10143-10162.

[4] R. A. Fisher, Statistical Methods and Scientific Inference, Oliver & Boyd, Edinburgh, 1956.

[5] Pierre Lafaye De Micheaux and Benoit Liquet, Understanding convergence concepts: a visual-minded and graphical simulation-based approach, Amer. Statist. 63(2) (2009), 173-178.

[6] Francois Le Gland, M. Valerie and V. D. Tran, Large sample asymptotics for ensemble Kalman filter, Institut National De Recherche En Informatique En Automatique, 2009.

[7] J. Li and Dongbin Xiu, On numerical properties of the ensemble Kalman filter for data assimilation, Comput. Methods Appl. Mech. Engrg. 197 (2008), 3574-3583.

[8] J. Mandel, L. Cobb and J. D. Beezley, On the convergence of the ensemble Kalman filter, University of Colorado Denver CCM Report 278, January 2009. http://www.arXiv.org/abs/0901.2951.

[9] M. G. Marasinghe, W. Q. Meeker, D. Cook and T. Shin, Using graphics and simulation to teach statistical concepts, Amer. Statist. 50 (1996), 342-351.

[10] D. C. Montgomery, Statistical Quality Control, A Modern Introduction, 6th Edition, John Wiley & Sons, Inc., 2009.

[11] M. Tan, Mathematical properties of ensemble Kalman filter, Dissertation of Faculty of The USC Graduate School University of California, 2011.