A deep learning-based ensemble filter for nonlinear data assimilation

Truong-Vinh Hoang1,* and Hermann G. Matthies2,++

1 Chair of Mathematics for Uncertainty Quantification, RWTH Aachen University, 52056 Aachen
2 Institute for Scientific Computing, Technische Universität Braunschweig, 38106 Braunschweig

This article presents a novel deep learning-based ensemble conditional mean filter (DL-EnCMF) for nonlinear data assimilation. The filter’s key component is the approximation of the conditional expectation (CE) using deep neural networks (DNNs). We implement the DL-EnCMF for tracking the states of the Lorenz-63 system. Numerical results show that the DL-EnCMF outperforms the ensemble Kalman filter (EnKF)—a common technique for data assimilation.

© 2021 The Authors Proceedings in Applied Mathematics & Mechanics published by Wiley-VCH GmbH

1 Introduction

Data assimilation aims to update states of a dynamical system by combining numerical models and observations, which can be sparse in space and time [1]. Due to the presence of uncertainty, the prediction of the system states is often modeled using probability distributions. When new measurement data become available, the forecast distribution of the states is updated by conditioning on the observations. For high-dimensional and nonlinear dynamical systems, an accurate representation of the conditional distribution comes at an extremely high computational cost. A common method for data assimilation with an acceptable computational budget is the EnKF. However, the EnKF is not appropriate for highly nonlinear dynamical systems as the filter uses linear approximations for the state evolution and the observation map.

The DL-EnCMF presented in this paper is a generalization of the EnKF for the nonlinear case. The main difficulty in implementing the EnCMF is the approximation of the conditional expectation (CE) of the forecast states. The DNN, which has significant advantages in representing complex functions between high-dimensional spaces [2], is applied to approximate the unknown measurement errors which are modeled as a RV denoted as $\xi_k$.

2 DL-based ensemble conditional expectation mean filter

In this section, we present the formulation of the EnCMF and its implementation using DNNs. Consider a dynamical system governed by the following stochastic ordinary differential equation:

$$\frac{d}{dt} Q(t) = f(Q(t)), \quad t > 0, \quad Q(t = 0) = Q_0,$$

where $Q(t)$ and $Q_0$ are $n$-dimensional random vectors (RVs) of the system states at time $t$ and $t = 0$ respectively, and $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Data assimilation aims to track the true states of the system $q(t)$ given observations $y_k$ at time $t_k$, $y_k = h(q_k) + \xi_k$, where $k = 1, 2, \ldots, K$, $y_k$ is an $m$-dimensional vector of the observations, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the observation map, and $\xi_k$ is the unknown measurement errors which are modeled as a RV denoted as $\Xi_k$. In practice, the observation time interval $\Delta t_{\text{obs}}$, assumed being constant ($\Delta t_{\text{obs}} = t_{k+1} - t_k$, for $k = 1, \ldots, K - 1$), is much larger than the time step used for the temporal discretization of the dynamical system. The CMF consists of two steps [3]:

- prediction step: $Q_k^f = Q_{k-1}^f + \phi(t_{k-1}, f(Q(t))) dt$, and
- analysis step: $Q_k^m = h(Q_k^f) + \Xi_k$.

where $Q_k^f$ and $Q_k^m$ are the forecast and the assimilated RVs, respectively, at the time $t_k$, $Y_k^f$ is the RV of predicted observations, $\Phi_{Q_k^f[Y_k^f]} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is the map of the conditional mean satisfying $\Phi_{Q_k^f[Y_k^f]}(Y_k^f) = \mathbb{E}(Q_k^f|Y_k^f)$, where $\mathbb{E}$ is the expectation operator. The EnCMF is the implementation of the CMF using the Monte Carlo method. Let \{\{q_k^{(1)}, \ldots, q_k^{(N)}\}\} and \{\{y_k^{(1)}, \ldots, y_k^{(N)}\}\} be sets of samples of the RVs $Q_k^f$ and $Y_k^f$, respectively. Using Eq. (2), we obtain the ensemble of the assimilated state: $q_k^{(i)} = \Phi_{Q_k^f[Y_k^f]}(Y_k^{(i)}) + \xi_k$, for $i = 1, \ldots, N$. We remark that the mean of the assimilated ensemble converges to the conditional expectation $\mathbb{E}(Q_k^f|y_k)$, i.e., $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^{N} q_k^{(j)} = \mathbb{E}(Q_k^f|y_k)$.

* Corresponding author: e-mail hoang@tu-activated-aachen.de
++ Corresponding author: e-mail wire@tu-bs.de
A method to identify the map \( \Phi_{Q_f|Y_f}^k \) is to solve the following minimization problem [3,4]:

\[
\Phi_{Q_f|Y_f}^k = \arg \min_G \mathbb{E}(||G(Y_k^f) - Q_k^f||^2).
\]  

(3)

Limiting the function \( \Phi_{Q_f|Y_f}^k \) to be linear, the En-CMF becomes the EnKF. Due to this linear approximation, for the cases that \( h \) is nonlinear, or \( Q_k^f \) and \( Y_k^f \) are not Gaussian RVs, the property that the assimilated ensemble’s mean converges to \( \mathbb{E}(Q_k^f|y_k) \) is no longer guaranteed.

Here, we approximate \( \Phi_{Q_f|Y_f}^k \) using DNNs. Let \( G_{NN} : \mathbb{R}^m \rightarrow \mathbb{R}^n \) be a DNN with hyper-parameters \( \Theta \)—weights and biases at each neural of the network. Using to the condition stated in Eq. (3), we can identify the parameters \( \Theta \) as:

\[
\Theta = \arg \min_{\Theta} L(\Theta),
\]

where \( L(\Theta) = \mathbb{E}(||G_{NN}(Y_k^f|\Theta) - Q_k^f||^2) \) is the loss function. Given the forecast ensembles, the loss function is approximated as \( L(\Theta) \approx \frac{1}{N} \sum_{j=1}^{N} ||G_{NN}(y_k^{f(j)}|\Theta) - f_k^j||^2 \). It is noted that variational control techniques can be applied to reduce the statistical error in estimating the loss \( L(\Theta) \).

**Numerical examples** We demonstrate the DL-EnCMF for tracking states of the Lorenz-63 system, which is described as:

\[
\frac{dq_1}{dt} = \sigma(q_2 - q_1), \quad \frac{dq_2}{dt} = q_1(\rho - q_3) - q_2, \quad \frac{dq_3}{dt} = q_1q_2 - \beta q_3,
\]

where \( \sigma = 10, \beta = 8/3, \) and \( \rho = 28 \). To simulate the system, we use the forward time integration scheme with a time step 0.05. The observation operator is modelled as

\[
y_k = q(t_k) + \xi_k,
\]

where \( \Xi_k \sim \mathcal{N}(0, 2I_3) \). The DL-EnCMF and the EnKF are implemented for 2000 assimilation steps using the same ensemble size, \( N = 1000 \). In the DL-EnCMF, a dense feed-forward neural network with two hidden layers, composed of 20 neurons, is employed to approximate the conditional mean. To evaluate filter performance, we use the average root mean square error (RMSE) defined as \( \text{rmse} = \left[ \frac{1}{N} \sum_{j=1}^{N} ||q_j^a - q_j(t_j)||^2 \right]^{1/2} \), where \( q_j^a \) is the assimilated ensemble mean. The obtained values of \( \text{rmse} \) are illustrated in Fig. 1. As it is observed from Fig. 1, for small observation time intervals, \( \Delta t_{\text{obs}} < 0.2 \), the DL-EnCMF and the EnKF has similar average RMSEs. However, for larger observation time intervals, the DL-EnCMF significantly outperforms the EnKF.

![Fig. 1: Comparison of EnKF and DL-EnCMF for tracking Lorenz 63 system in terms of average RMSE.](image)

**3 Conclusion**

This article presents the novel DL-EnCMF for nonlinear data assimilation. The filter’s core element is the approximation of the CE using DNNs. Asserting the filter’s performance for high-dimensional dynamical systems is desirable for future work.

**Acknowledgements** Open access funding enabled and organized by Projekt DEAL.

Correction added on 19 February 2021, after first online publication: Hermann G. Matthies was designated as corresponding author.
References

[1] K. Law, A. Stuart, and K. Zygalakis, Data assimilation (Springer, 2015).
[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning (MIT press, 2016).
[3] H. G. Matthies, E. Zander, B. V. Rosić, and A., Litvinenko, Adv. Model. and Simul. in Eng. Sci. (2016) 3:24.
[4] A. Bobrowski. Functional analysis for probability and stochastic processes: an introduction. (Cambridge University Press, 2005).