# Uniform Propagation of Chaos in Ensemble Kalman Filters

Danilo Jr Dela Cruz
Mathematics of Random Systems
University of Oxford

21st January 2024

## Abstract

The Ensemble Kalman Filter (EnKF) enables us to perform inference in high dimensional settings by approximating the covariance matrix with an ensemble average. The approximation error can propagate leading to "catastrophic filter divergence". Del Moral and Tugaut [2016] provide a rigorous analysis of this error by interpreting EnKF as a McKean-Vlasov interacting particle system. We evaluate their results and conjecture that an intersection with the field of Dynamical Systems will yield more pertinent results.

## 1 Introduction

In Stochastic Filtering, we have a (unobserved) signal process $X_t \in \mathbb{R}^{r_1}$ and observation process $Y_t \in \mathbb{R}^{r_2}$. Through assimilation of observations and an understanding of their dynamics, we can improve our estimates for $X_t$. This is critical for numerical weather predictions as chaotic systems are prevalent and necessitate high accuracy.

We will assume that $X_t, Y_t$ are governed by linear diffusions. I.e. for suitably sized matrices $A, C$, symmetric positive definite matrices $R_1, R_2$, vectors $a, c$, independent standard brownian motions $W_t, V_t$ and an initial condition $X_0 \sim \mathcal{N}(\mu_0, P_0)$ independent of $W_t, V_t$, we define

$$dX_t = (AX_t + a)dt + R_1^{1/2} dW_t \tag{1}$$
$$dY_t = (CX_t + c)dt + R_2^{1/2} dV_t. \tag{2}$$

Let $\mathcal{F}_t = \sigma(Y_s : s \leq t)$ be the natural filtration of the observation process. Our goal is to find

$$\widehat{X}_t = \mathbb{E}\left[X_t \,|\mathcal{F}_t\right] \tag{3}$$

In this context of the linear dynamics, the solution is given by the Kalman-Bucy Equations 4

$$d\widehat{X}_t = d\tilde{X}_t + K_t dI_t \tag{4}$$
$$d\widetilde{X}_t = (A\widehat{X}_t + a)dt \tag{5}$$
$$dI_t = dY_t - (C\widehat{X}_t + c)dt. \tag{6}$$

At its core, we adjust our prediction 5 (what we get by evolving under the dynamics) by the innovation process 6 (which represents the difference between the predicted and realised observation). The level to which we adjust is determined by the Kalman Gain 7 which is a ratio between the uncertainty $P_t$ in the initial prediction $\widetilde{X}_t$ and the noise $R_2$ in the observation $Y_t$.

$$K_t = P_t C^T R_2^{-1} \tag{7}$$

This requires us to track the covariance matrix $P_t$ of $\widehat{X}_t$ whose evolution can be described by a Riccati Equation 8.

$$\partial_t P_t = \mathrm{Ricc}(P_t) \tag{8}$$
$$= AP_t + P_t A^T - P_t S P_t + R_1 \tag{9}$$
$$S = C^T R_2^{-1} C \tag{10}$$

However, updating $P_t$ becomes intractable in high dimensional settings when $r_1, r_2$ are large. In weather forecasting, it is typical to encounter applications with $r_1 \geq 10^7, r_2 \geq 10^5$ Katzfuss et al. [2016]. The Ensemble Kalman Filter can alleviate this cost by a factor of $O(N/r_1)$ where $N$ is the size of the ensemble. See Appendix A.

### 1.1 Ensemble Kalman Filter

By having an ensemble of particles $\xi_t^i$ which mimic $\widehat{X}_t$, we can use their sample mean $m_t$ to estimate

$\widehat{X}_t$ and use the sample covariance $p_t$ to estimate $P_t$. Formally, let $(\overline{W}^i_t, \overline{V}^i_t, \xi^i_0)_{i=1}^N$ be $N$ independent copies of $(W_t, V_t, X_0)$. Define

$$d\xi^i_t = dS^i_t + p_t C^T R_2^{-1}\left[dY_t - dO^i_t\right] \tag{11}$$

$$dS^i_t = (A\xi^i_t + a)dt + R_1^{1/2} d\overline{W}^i_t \tag{12}$$

$$dO^i_t = (C\xi^i_t + c)dt + R_2^{1/2} d\overline{V}^i_t \tag{13}$$

$$p_t = \frac{1}{N-1}\sum_{i\leq i\leq N}(\xi^i_t - m_t)(\xi^i_t - m_t)^T \tag{14}$$

$$m_t = \frac{1}{N}\sum_{1\leq i\leq N}\xi^i_t. \tag{15}$$

Given the ubiquity of EnKF in applications, it is important to understand the error of these estimates. In particular, these errors can propagate as $\xi^i_t$ are updated using $p_t$. The next section states and evaluates the results from Del Moral and Tugaut [2016] which address these errors.

# 2 Propagation of Chaos

Define the following conditions:

- **Stability Condition** - Let $\mu(A) < 0$ [1].
- **Observability Condition** - Let $S = \rho(S)\,\mathrm{Id}$ be a scaled identity matrix for $\rho(S) > 0$.

Under these conditions, Theorem 3.6 and Corollary 3.7 in Del Moral and Tugaut [2016] provide uniform in time error bounds for $(m_t, p_t)$.

**Theorem 2.1** (Uniform Propagation of Chaos)**.** *Under the observability and stability condition, for any $p \geq 1$ and sufficiently large $N > 1 + 2r_1(3p-1)$*

$$\sup_{t\geq 0}\mathbb{E}\left[\left\|m_t - \widehat{X}_t\right\|_2^p\right]^{1/p} \leq c(p)/\sqrt{N} \tag{16}$$

$$\sup_{t\geq 0}\mathbb{E}\left[\|p_t - P_t\|_F^p\right]^{1/p} \leq c(p)/\sqrt{N}\ ^{[2]} \tag{17}$$

*for some constant $c(p)$ depending on $p$.*

The authors note that without these conditions, the statement is valid if we drop the supremum over time and our constant factor $c_t(p)$ would depend on time. This is already interesting, as we may only be concerned with the error for our prediction at a given point in time.

In order to evaluate these assumptions, we begin by the motivation provided in Section 4.

## 2.1 Motivation for Assumptions

For simplicity, suppose a steady state $P$ exists for equation 8 and set $P_0 = P = P_t$. Let $Z_t$ represent an ensemble member which uses $p_t = P + Q_t$, $Q_t$ representing the sampling error. In Section 4.1 they show that the error can be expressed in terms of $E_t$.

$$Z_t - X_t = E_t(Z_0 - X_0) + \int_0^t E_t E_s^{-1} d\mathcal{W}_s \tag{18}$$

where $\mathcal{W}_s$ is some diffusion and

$$E_t = \exp\left(\oint_0^t (A - (P + Q_s)S)ds\right). \tag{19}$$

Clearly if $E_t$ is small then the error $Z_t - X_t$ is small. A semigroup estimate yields

$$\|E_t\|_2 \leq \exp\left(\int_0^t \mu(A - (P + Q_s)S)ds\right). \tag{20}$$

In order to ensure $E_t$ decays, we can assert the logarithmic normm $\mu(A - (P + Q_s)S) < -\delta$ which allows for uniform error estimates. By properties of the logarithmic norm

$$\begin{array}{rcl} \mu(A - (P + Q_s)S) & \leq & \mu(A) + \mu(-(P + Q_s)S) \\ \text{(Observability)} & = & \mu(A) + \rho(S)\mu(-(P + Q_s)) \\ (P, Q_s \in \mathbb{S}^+) & \leq & \mu(A) \\ \text{(Stability)} & < & 0. \end{array}$$

The conditions lets us use $-\delta = \mu(A)$. Next we describe what these assumptions correspond to.

## 2.2 Interpretation of Assumptions

These are strong assumptions and it limits their applicability. Indeed the stability condition is obtained by ensuring the upper bound for $E_t$ decays. However, we should expect our errors to increase with time and instead concern ourselves with the rate at which the error grows. Perhaps asking for a supremum over $[0, \infty)$ as opposed to $[0, T]$ is unreasonable and would only work for special processes.

[1] Recall that the logarithmic norm $\mu(A)$ is an upper bound on the real part of the eigenvalues of $A$. This means that the signal process is an Ornstein-Uhlenbeck process and hence stable.

[2] $F$ indicates Frobenius Norm.

### 2.2.1 Stability Condition

$\mu(A) < 0$ means that the signal is an OU process. I.e. it hovers around its mean $a$. This excludes many types of signal processes.

Furthermore, the authors remark that this is a necessary condition for uniform estimates if $R_1 \neq 0$ and $C = 0$. However, $C = 0$ means the observation process is noise which does not contain any information about the signal process. Conceivably, $\mu(A) > 0$ but the observation process is able to correct our prediction and avoid filter divergence.

### 2.2.2 Observability Condition

As described in Section 4.2 if $r_1 = r_2$ and $C$ is invertible, the $S$ corresponding to

$$\mathcal{X}_t = (R_2^{-1/2}C)X_t, \quad \mathcal{Y}_t = R_2^{-1/2}Y_t \tag{21}$$

satisfies this property. These correspond to the class of full observation sensors. Intuitively, the observation process is a perturbed linear transformation of the signal and the above transformation sends $X_t$ into the same basis as $Y_t$. Models like the Lorentz-96 filtering problems are full-observation but often $r_2 < r_1$ as some states are unobservable. The authors conjecture that this condition can be relaxed.

Finally, we discuss how requiring sufficiently large $N$ impacts the pertinence of Theorem 2.1.

# 3 Pertinence

In the introduction, we noted that Ensemble Kalman Filter arose in order to alleviate the computational burden of updating the covariance matrix $P_t$. In particular, the cost is multiplied by $O(N/r_1)$ which is benificial if $N < r_1$. However, Theorem 2.1 requires $N > 1 + 2r_1(3p - 1)$ for some $p \geq 1$ which would render EnKF more expensive.

To get some context, we consider the example of Numerical Weather Prediction from Chattopadhyay et al. [2023]. We have 10 quantities of interest such as temperature, wind velocity and pressure which we want to model over a 3D lattice with a resolution of 100 points for each dimension. This yields $r_1 = 10 \cdot 10^2 \cdot 10^2 \cdot 10^2 = 10^7$. The ensemble sizes used in practice is around $N = 50$. This yields a reduction in cost by a factor of $10^6$ which makes the problem tractable. However, it is far from the guarantees of the theorem and we are unable to make statements about its stability.

Indeed, it is empirically observed that small ensemble sizes without modifications lead to filter divergence. To prevent this, ad-hoc strategies are applied, see Katzfuss et al. [2016].

- **Localisation** - Strong correlations between points that are far apart in the lattice are likely to be spurious and so we smoothen $p_t$.
- **Variance Inflation** - The sample covariance matrix often underestimates the true covariance matrix and our ensemble is unable to be corrected by the innovation process. To mitigate against this, we artificially inflate $p_t$.

## 3.1 Model Dynamics

As discussed in section 2, the assumptions are too general. In Ng et al. [2011], they study how the dynamics of the signal process plays a role in filter divergence. They show that the ensemble aligns itself with the subspace spanned by unstable Lyapunov vectors and divergence can be avoided if and only if this subspace is fully spanned.

This demonstrates how insights from Dynamical Systems can lead to more sophisticated conditions. Perhaps integrating these insights with the rigorous analysis of Del Moral and Tugaut [2016] will yield pertinent results.

# 4 Conclusion

Del Moral and Tugaut [2016] have pioneered convergence analysis of the Ensemble Kalman Filter by interpreting it as a McKean-Vlasov interacting particle system. In the process, they have constructed a framework which will enable future research.

Currently, the assumptions are too general and Theorem 2.1 is impertinent to practitioners. Perhaps, integrating insights from dynamical systems and studying the ad-hoc mechanisms used in practice will lead to influential results which will inform practitioners how to employ EnKF.

# A Computational Cost

To estimate the cost of the Kalman Filter we consider an Explicit Euler Discretisation. I.e.

$$X_{t+1} = X_t + dX_t \tag{22}$$
$$P_{t+1} = P_t + \partial_t P_t. \tag{23}$$

For EnKF, there are variations discussed in Section 4 of Katzfuss et al. [2016] such as serial filtering, which alleviate the storage burden in high dimensional settings. However, we will focus on the basic EnKF which is described in equation 11.

Applying $n$ iterations and an ensemble size of $N$, we can show the costs given in Table 1. For the Kalman Filter, the cost $O(nr_1(r_1+r_2)^2)$ is dominated by the covariance update whereas EnKF has a cost dominated by the state update $O(nN(r_1+r_2)^2)$. The ratio of their costs is $O(N/r_1)$.

| Cost | Euler | EnKF |
|---|---|---|
| State [3] | $O(n(r_1+r_2)^2)$ | $O(nN(r_1+r_2)^2)$ |
| Covariance | $O(nr_1(r_1+r_2)^2)$ | $O(nNr_1^2)$ |
| Leading | $O(nr_1(r_1+r_2)^2)$ | $O(nN(r_1+r_2)^2)$ |
| Storage | $O(r_1^2)$ | $O(r_1^2)$ |

Table 1: Cost comparison over $n$ iteration of Kalman Filter vs Ensemble Kalman Filter with ensemble size $N$. Leading refers to the max computational cost. Storage only refers to the additional required to store $p_t$.

[3] To evaluate the cost of $ABx$, where $A, B$ are matrices, one should always avoid matrix-matrix multiplications and do recursive matrix-vector multiplication. This is because $A(Bx)$ is $O(r^2)$ whereas $(AB)x$ which is $O(r^3 + r^2)$.

# References


Ashesh Chattopadhyay, Ebrahim Nabizadeh, Eviatar Bach, and Pedram Hassanzadeh. Deep learning-enhanced ensemble-based data assimilation for high-dimensional nonlinear dynamical systems. *Journal of Computational Physics*, 477:111918, March 2023. ISSN 00219991. doi: 10.1016/j.jcp.2023.111918.

Pierre Del Moral and Julian Tugaut. On the stability and the uniform propagation of chaos properties of Ensemble Kalman-Bucy filters, October 2016.

Matthias Katzfuss, Jonathan R. Stroud, and Christopher K. Wikle. Understanding the Ensemble Kalman Filter. *The American Statistician*, 70(4):350–357, October 2016. ISSN 0003-1305, 1537-2731. doi: 10.1080/00031305.2016.1141709.

Gene-Hua Crystal Ng, Dennis McLaughlin, Dara Entekhabi, and Adel Ahanin. The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A: Dynamic Meteorology and Oceanography*, 63(5):958, January 2011. ISSN 1600-0870. doi: 10.1111/j.1600-0870.2011.00539.x.