

# Optimality vs Stability Trade-off in Ensemble Kalman Filters

Amirhossein Taghvaei\* Prashant G. Mehta\*\* Tryphon T. Georgiou\*\*\*

\* Department of Aeronautics and Astronautics, University of Washington, Seattle CA (e-mail: amirtag@uw.edu).

\*\* Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, IL (e-mail: mehtapg@illinois.edu)

\*\*\* Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA (e-mail: tryphon@uci.edu)

**Abstract:** This paper is concerned with optimality and stability analysis of a family of ensemble Kalman filter (EnKF) algorithms. EnKF is commonly used as an alternative to the Kalman filter for high-dimensional problems, where storing the covariance matrix is computationally expensive. The algorithm consists of an ensemble of interacting particles driven by a feedback control law. The control law is designed such that, in the linear Gaussian setting and asymptotic limit of infinitely many particles, the mean and covariance of the particles follow the exact mean and covariance of the Kalman filter. The problem of finding a control law that is exact does not have a unique solution, reminiscent of the problem of finding a transport map between two distributions. A unique control law can be identified by introducing control cost functions, that are motivated by the optimal transportation problem or Schrödinger bridge problem. The objective of this paper is to study the relationship between optimality and long-term stability of a family of exact control laws. Remarkably, the control law that is optimal in the optimal transportation sense leads to an EnKF algorithm that is not stable.

*Keywords:* filtering, mean-field control, optimal transportation

## 1. INTRODUCTION

Consider the linear system

$$dX_t = AX_t + \sigma_B dB_t, \quad X_0 \sim N(m_0, \Sigma_0) \quad (1a)$$

$$dZ_t = HX_t dt + dW_t, \quad (1b)$$

where  $X_t \in \mathbb{R}^d$  is the state of the system at time  $t$ ,  $Z_t \in \mathbb{R}^m$  is the observation process,  $B_t \in \mathbb{R}^d$  and  $W_t \in \mathbb{R}^m$  are mutually independent standard Brownian motions, and  $A$ ,  $H$ ,  $\sigma_B$  are matrices of appropriate dimensions. The distribution of the initial state  $X_0$  is Gaussian  $N(m_0, \Sigma_0)$  with mean  $m_0$  and covariance  $\Sigma_0$ .

The filtering problem is to compute the *posterior distribution* of  $X_t$  conditioned on the filtration generated by the history of observations  $\mathcal{Z}_t := \sigma(Z_s; 0 \leq s \leq t)$ . For the linear system (1), the filtering problem admits an explicit solution: the posterior distribution is Gaussian  $N(m_t, \Sigma_t)$  with mean and covariance governed by the Kalman-Bucy filter equations [Kalman and Bucy 1961]:

$$dm_t = Am_t dt + K_t(dZ_t - Hm_t dt) =: \mathcal{T}_t(m_t, \Sigma_t), \quad (2a)$$

$$\frac{d\Sigma_t}{dt} = A\Sigma + \Sigma A^\top + \Sigma_B - \Sigma H^\top H \Sigma =: \text{Ricc}(\Sigma_t), \quad (2b)$$

where  $K_t := \Sigma_t H^\top$  is the Kalman gain and  $\Sigma_B := \sigma_B \sigma_B^\top$ . The notation  $\mathcal{T}_t(\cdot, \cdot)$  and  $\text{Ricc}(\cdot)$  is used to identify the update law for the mean and covariance respectively.

Ensemble Kalman filter (EnKF) is a Monte-Carlo-based numerical algorithm that is designed to approximate the solution to the filtering problem [Evensen 1994, Whitaker and Hamill 2002, Reich 2011, Bergemann and Reich 2012]. EnKF is widely used in in applications (such as weather prediction) where the state

dimension  $d$  is very high; cf., [Bergemann and Reich 2012, Houtekamer and Mitchell 2001]. The high-dimension of the state-space provides a significant computational challenge *even* in linear Gaussian settings. For such problems, an EnKF implementation may require less computational resources (memory and FLOPS) than a Kalman filter [Houtekamer and Mitchell 2001, Evensen 2006].

The EnKF algorithm is designed in two steps:

- (i) a controlled stochastic process  $\bar{X}_t$  is constructed, whose conditional distribution given  $\mathcal{Z}_t$  is equal to the posterior distribution of  $X_t$ :

$$\text{(exactness)} \quad P(X_t \in \cdot | \mathcal{Z}_t) = P(\bar{X}_t \in \cdot | \mathcal{Z}_t). \quad (3)$$

- (ii) an ensemble of  $N$  stochastic processes  $\{X_t^i\}_{i=1}^N$  is simulated to empirically approximate the distribution of  $\bar{X}_t$ :

$$P(\bar{X}_t \in \cdot | \mathcal{Z}_t) \approx \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}(\cdot). \quad (4)$$

The process  $\bar{X}_t$  is referred to as *mean-field process* and the stochastic processes  $\{X_t^i\}_{i=1}^N$  are referred to as *particles*. The property (3) is referred to as *exactness*.

The motivation to study the EnKF algorithm is two-fold:

- (i) As mentioned above, EnKF algorithm is computationally efficient compared to Kalman filter algorithm for high-dimensional problems [Bergemann and Reich 2012]. The computational cost of EnKF scales as  $O(Nd)$ , whereas it scales as  $O(d^2)$  for Kalman filter.

Table 1. Description of three established forms of EnKF.  $G_t, r_t, q_t$  are parameters of the mean-field process update law (8). Stability rate and approximation error are described in Prop. 2 and 3 respectively.

Algorithm	Acronym	$G_t$	$r_t$	$q_t$	Stability rate	steady-state error
EnKF with perturbed observation [Reich 2011, Eq. (26)]	P-EnKF	$A - \bar{\Sigma}_t H^2$	$\sigma_B$	$\bar{\Sigma}_t H$	$\lambda_0$	$\propto N^{-1}(\Sigma_B + H^2)$
Square-root EnKF [Bergemann and Reich 2012, Eq. (3.3)]	S-EnKF	$A - \frac{1}{2}\bar{\Sigma}_t H^2$	$\sigma_B$	0	$\frac{1}{2}(\lambda_0 - A)$	$\propto N^{-1}\Sigma_B$
Deterministic EnKF [Taghvaei and Mehta 2016]	D-EnKF	$A - \bar{\Sigma}_t H^2 + \frac{1}{2}\Sigma_B \bar{\Sigma}_t^{-1}$	0	0	0	0

(ii) The two-step design procedure can be generalized to *non-linear* and *non-Gaussian* setting. The result is the feedback particle filter (FPF) algorithm [Yang et al. 2013, 2016]. Therefore, EnKF can be considered as special case of FPF, and understanding EnKF is insightful for understanding the FPF algorithm.

Here is the outline and summary of the paper:

(i) The problem of constructing a mean-field process  $\bar{X}_t$ , such that exactness property (3) is satisfied, is addressed in Section 2. It is shown that exact mean-field process is not unique. The family of exact processes is identified in Prop. 1. Three members of the family, that correspond to three established forms of EnKF algorithm, are presented in Table 1.

(ii) The stability of the mean-field process is studied in Section 3. It is shown that, the mean-field processes exhibit different stability behaviour. In particular, the deterministic EnKF, that is constructed from limit of incremental optimal transportation maps, is not stable. While, other forms EnKF that involve stochastic terms, are stable.

(iii) Finally, the particle system and the analysis of the approximation error (4) is presented in Section 4. It is observed that the EnKF algorithms that were stable, exhibit larger approximation error, because of the presence of stochastic terms.

This paper builds on the growing literature on the design and the analysis of the EnKF algorithms. Three forms of the EnKF algorithm are of importance: EnKF with perturbed observation [Evensen 1994, Reich 2011], square-root form of EnKF [Whitaker and Hamill 2002, Bergemann and Reich 2012] which is the same as FPF algorithm with constant gain approximation [Yang et al. 2016], and deterministic EnKF [Taghvaei and Mehta 2016, 2020]. This paper is also related to the stability and the error analysis of the EnKF algorithm in the discrete-time setting [Le Gland et al. 2009, Mandel et al. 2011, Tong et al. 2016, Kelly et al. 2014, Kwiatkowski and Mandel 2015], and the continuous-time setting [Del Moral and Tugaut 2018, de Wiljes et al. 2018, Del Moral et al. 2017, Bishop and Del Moral 2018].

**Assumption I:** It is assumed that the processes  $X_t$  and  $Z_t$  are scalar, i.e.  $d = m = 1$ . This assumption is made to simplify the exposition. The possible extension to vector-valued case is briefly discussed as a remark in each Section.

## 2. CONSTRUCTION OF MEAN-FIELD PROCESS

In order to construct an exact mean-field process  $\bar{X}_t$ , consider the following sde

$$d\bar{X}_t = \mathcal{U}_t(\bar{X}_t) dt + \mathcal{V}_t(\bar{X}_t) dZ_t + \mathcal{R}_t(\bar{X}_t) d\bar{B}_t + \mathcal{Q}_t(\bar{X}_t) d\bar{W}_t, \quad \bar{X}_0 \sim \bar{\pi}_0, \quad (5)$$

driven by control laws  $\mathcal{U}_t(\cdot), \mathcal{V}_t(\cdot), \mathcal{R}_t(\cdot), \mathcal{Q}_t(\cdot)$ , where  $\bar{B}_t$  and  $\bar{W}_t$  are independent copies of  $B_t$  and  $W_t$  of (1), and  $\bar{\pi}_0$  is the initial distribution. The following result characterizes

control laws that lead to a mean-field process  $\bar{X}_t$  with exactness property (3).

*Proposition 1.* The mean-field process (5) satisfies the exactness property (3) if the initial distribution  $\bar{\pi}_0$  is Gaussian  $N(m_0, \Sigma_0)$ , and

$$\mathcal{U}_t(x) = (A - \bar{\Sigma}_t H^2)\bar{m}_t + G_t(x - \bar{m}_t), \quad (6a)$$

$$\mathcal{V}_t(x) = \bar{\Sigma}_t H, \quad (6b)$$

$$\mathcal{R}_t(x) = r_t, \quad \mathcal{Q}_t(x) = q_t, \quad (6c)$$

where  $\bar{m}_t = \mathbb{E}[\bar{X}_t | \mathcal{Z}_t]$ ,  $\bar{\Sigma}_t = \mathbb{E}[(\bar{X}_t - \bar{m}_t)^2 | \mathcal{Z}_t]$ , and  $G_t, r_t, q_t \in \mathbb{R}$  satisfy

$$2G_t \bar{\Sigma}_t + r_t^2 + q_t^2 = \text{Ric}(\bar{\Sigma}_t). \quad (7)$$

Using the form of control laws in (6), the sde (5) is

$$d\bar{X}_t = \mathcal{T}_t(\bar{m}_t, \bar{\Sigma}_t) + G_t(X_t^i - \bar{m}_t) dt + r_t d\bar{B}_t + q_t d\bar{W}_t, \quad \bar{X}_0 \sim \bar{\pi}_0 \quad (8)$$

where  $\mathcal{T}_t(m, \Sigma)$  is defined in (2). The fact that  $\bar{X}_t$  satisfies the exactness property (3) is observed by noting that: (i)  $\bar{m}_t = m_t$  and  $\bar{\Sigma}_t = \Sigma_t$ , because the time-evolution of  $\bar{m}_t$  and  $\bar{\Sigma}_t$  is identical to the Kalman filter equations (2); (ii) the distribution of  $\bar{X}_t$  is Gaussian because the sde (8) is linear upon replacing  $\bar{m}_t = m_t$  and  $\bar{\Sigma}_t = \Sigma_t$  and the initial distribution  $\pi_0$  is Gaussian.

There are three established choices for the parameters  $G_t, r_t, q_t$  that are tabulated in Table 1, where each row corresponds to a specific form of EnKF algorithm.

*Remark 1.* The form of the sde (5) may not be general enough to capture all possible stochastic processes  $\bar{X}_t$  that achieve the exactness property (3). The particular form of sde (5) is motivated by its appealing control theoretic form, where all the control terms are assumed to be of feedback form. For a more general prescription of exact mean-field processes, see Abedi and Surace [2019].

*Remark 2.* For the vector-valued case,  $\mathcal{U}(x)$  involves additional divergence free term (e.g.  $\psi(x) = \Omega \Sigma^{-1}(x - \bar{m}_t)$  with skew-symmetric matrix  $\Omega$ ) that does not effect the distribution [Taghvaei and Mehta 2020].

*Remark 3.* The control law governing the deterministic EnKF in Table 1 is optimal with respect to control cost associated with optimal transportation problem. In particular, it is obtained as the continuous-time limit of infinitesimal optimal transportation maps between the Gaussian distributions that are given by the Kalman filter [Taghvaei and Mehta 2020]. Also, the control law governing the square-root EnKF in Table 1 is optimal with respect to a control cost associated with the Schrödinger bridge problem with prior dynamics given by (1) [Chen et al. 2016].

## 3. STABILITY OF THE MEAN-FIELD PROCESS

Let  $\pi_t$  denote the exact posterior distribution given by Kalman-Bucy filter (2), and  $\bar{\pi}_t$  denote the distribution of the mean-field process given by (8). Proposition 1 informs that the exactness

condition  $\bar{\pi}_t = \pi_t$  is satisfied if the initial distribution  $\bar{\pi}_0 = \pi_0$  and (7) holds. The objective is to study the error between  $\bar{\pi}_t$  and  $\pi_t$  if the initial distributions  $\bar{\pi}_0$  and  $\pi_0$  are not equal.

The convergence analysis is carried out by analyzing the convergence of  $\bar{\pi}_t$  to the Gaussian distribution  $\tilde{\pi}_t = N(\bar{m}_t, \bar{\Sigma}_t)$  with the same mean and variance as the mean-field process, and the convergence of the mean and variance to the mean and variance given by the Kalman-Bucy filter equations. The convergence result is presented in the following proposition. We use 2-Wasserstein metric [Villani 2003], denoted by  $\mathcal{W}_2(\cdot, \cdot)$ , in order to measure the error between distributions, and we make the following assumption.

**Assumption II:** The linear system (1) is controllable and observable. In the scalar case, this amounts to  $\sigma_B \neq 0$  and  $H \neq 0$ .

*Proposition 2.* Consider the mean-field process (8) under assumption (I)-(II). Then,

(i) The mean and the variance of the mean-field process converge to the exact mean and variance given by the Kalman-Bucy filter (2):

$$\begin{aligned} \mathbb{E}[|\bar{m}_t - m_t|^2] &\leq (\text{const.})e^{-2\lambda_0 t} (|\bar{m}_0 - m_0|^2 + |\bar{\Sigma}_0 - \Sigma_0|^2), \\ |\bar{\Sigma}_t - \Sigma_t| &\leq (\text{const.})e^{-2\lambda_0 t} |\bar{\Sigma}_0 - \Sigma_0|, \end{aligned} \quad (9)$$

where  $\lambda_0 = (A^2 + H^2 \Sigma_B)^{\frac{1}{2}}$ .

(ii) The error between the mean-field distribution  $\bar{\pi}_t$  and the Gaussian distribution  $\tilde{\pi}_t = N(\bar{m}_t, \bar{\Sigma}_t)$  is bounded by:

$$\mathcal{W}_2(\bar{\pi}_t, \tilde{\pi}_t) \leq e^{\int_0^t G_s ds} \mathcal{W}_2(\bar{\pi}_0, \tilde{\pi}_0) \quad (10)$$

(iii) Combining part (i) and (ii), the total error between the mean-field distribution and the exact filter is bounded by

$$\mathbb{E}[\mathcal{W}_2(\pi_t, \bar{\pi}_t)] \leq C e^{-\lambda_0 t} \mathcal{W}_2(\tilde{\pi}_0, \pi_0) + e^{\int_0^t G_s ds} \mathcal{W}_2(\bar{\pi}_0, \tilde{\pi}_0) \quad (11)$$

where  $C$  is a constant independent of time  $t$ .

The convergence result (11) decomposes the error into two terms. The first term in the error is due to the incorrect specification of the initial mean and variance. In particular,

$$\mathcal{W}_2(\tilde{\pi}_0, \pi_0)^2 = (\bar{m}_0 - m_0)^2 + (\sqrt{\bar{\Sigma}_0} - \sqrt{\Sigma_0})^2.$$

The contribution from this term converges to zero with an exponential rate, independent of the choice made for  $G_t$ ,  $q_t$ ,  $r_t$  in (8) as long as (7) holds. The bound follows from the stability of the Kalman-Bucy filter which holds under controllability and observability of the linear system [Ocone and Pardoux 1996].

The second term in the error is due to the fact that the initial distribution is not Gaussian. It is controlled by the stability of the mean-field process (8) which, in contrast to the first term in the error, depends on the choice for  $G_t$ . For three choices of  $G_t$ , determined by the three forms of the EnKF tabulated in Table 1, the following holds

$$\text{(P-EnKF)} \quad e^{\int_0^t G_s ds} \leq (\text{const.})e^{-\lambda_0 t}, \quad (12a)$$

$$\text{(S-EnKF)} \quad e^{\int_0^t G_s ds} \leq (\text{const.})e^{-\lambda_1 t}, \quad (12b)$$

$$\text{(D-EnKF)} \quad e^{\int_0^t G_s ds} = \sqrt{\frac{\bar{\Sigma}_t}{\bar{\Sigma}_0}}, \quad (12c)$$

where  $\lambda_1 = \frac{\lambda_0 - A}{2}$ . Following conclusions are in order:

- (i) Both P-EnKF and S-EnKF are stable, i.e. the error converges to zero as  $t \rightarrow \infty$ . Moreover,  $\lambda_0 > \lambda_1 > 0$  (because  $\lambda_0 = (A^2 + H^2 \sigma_B^2)^{\frac{1}{2}} > |A|$ ). Therefore, the convergence rate of P-EnKF is strictly larger than the convergence rate of S-EnKF.
- (ii) D-EnKF is not stable. If the initial distribution is non-Gaussian, it remains non-Gaussian. In fact, one can establish the asymptotic lower-bound

$$\liminf_{t \rightarrow \infty} \mathcal{W}_2(\bar{\pi}_t, \pi_t) \geq (\text{const.}) \mathcal{W}_2(\bar{\pi}_0, \tilde{\pi}_0)$$

implying that the error remains positive if the initial distribution is not Gaussian.

*Remark 4.* The result (11) can be extended to vector-valued case by replacing  $\exp(\int_0^t G_s ds)$  with the state transition matrix  $\Phi_t$  defined as the solution to  $\frac{d}{dt} \Phi_t = G_t \Phi_t$  with  $\Phi_0 = I$ .

#### 4. PARTICLE SYSTEM AND ERROR ANALYSIS

The system of particles  $\{X_t^i\}_{i=1}^N$  is constructed from the mean-field process (8) by empirically approximating the mean and covariance:

$$\begin{aligned} dX_t^i &= \mathcal{T}_t(m_t^{(N)}, \Sigma_t^{(N)}) + G_t(X_t^i - m_t^{(N)}) dt + r_t dB_t^i \\ &\quad + q_t dW_t^i, \quad X_0^i = x_0^i, \quad \text{for } i = 1, \dots, N \end{aligned} \quad (13)$$

where  $m_t^{(N)} = N^{-1} \sum_{i=1}^N X_t^i$  and  $\Sigma_t^{(N)} = (N-1)^{-1} \sum_{i=1}^N (X_t^i - m_t^{(N)})^2$  are the empirical mean and the empirical covariance of the particles respectively, and  $\{B_t^i\}_{i=1}^N$  and  $\{W_t^i\}_{i=1}^N$  are independent copies of  $B_t$  and  $W_t$ . The initial state of the particles are denoted by  $\{x_0^i\}_{i=1}^N$ .

The objective is to analyze the error between the mean-field distribution  $\bar{\pi}_t$  and the empirical distribution of the particles

$$\pi_t^{(N)}(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}(\cdot),$$

where  $\delta_x$  is the Dirac delta distribution located at  $x$ . The result is presented for the convergence of the empirical covariance to the mean-field covariance under the following assumption:

**Assumption III:**  $\sup_{t \geq 0} \mathbb{E}[(r_t^2 + q_t^2) \Sigma_t^{(N)}] = M < \infty$ .

*Proposition 3.* Consider the particle system (13) and the mean-field process (8) under Assumptions (I)-(II)-(III). Then, the error between the empirical variance and the mean-field variance is bounded according to

$$\mathbb{E}[|\Sigma_t^{(N)} - \bar{\Sigma}_t|^2] \leq (\text{const.}) \left[ \frac{M}{N} + e^{-2\lambda_0 t} \mathbb{E}[|\Sigma_0^{(N)} - \bar{\Sigma}_0|^2] \right] \quad (14)$$

for  $N > \frac{16H^4 M}{(\lambda_0 - A)^2 \lambda_0}$ .

This result forms the basis for the convergence of the empirical distribution to the mean-field distribution. However, the analysis is more involved and the subject of ongoing work. We conjecture a result of the form

$$\begin{aligned} d(\pi_t^{(N)}, \bar{\pi}_t) &\lesssim \frac{M}{N} + e^{\int_0^t G_s ds} d(\pi_0^{(N)}, \bar{\pi}_0) \\ &\quad + e^{-2\lambda_0 t} (\mathbb{E}[|m_0^{(N)} - \bar{m}_0|^2 + |\Sigma_0^{(N)} - \bar{\Sigma}_0|^2]) \end{aligned} \quad (15)$$

where  $d(\mu, \nu) = \sup_{\|\nabla f\|_\infty < 1} \mathbb{E}[\mu(f) - \nu(f)]^2$  is a metric between two (possibly random) probability measure  $\mu$  and  $\nu$ , and  $\mu(f) := \int f d\mu$ . The convergence analysis is carried out

in the literature, for the three forms of the EnKF in Table 1, under strong assumption that the linear system is stable, i.e.  $A < 0$  [Del Moral and Tugaut 2018, Bishop and Del Moral 2018, Taghvaei 2019] (see Remark 5).

The error result (14) involves two terms. The second term is due to the error in the initial variance. It converges to zero as  $t \rightarrow \infty$  for any choice of  $G_t, r_t, q_t$ , as long as (7) holds, due to the stability of the Kalman-Bucy filter. The second term is due to the stochastic terms present in the particle system. It is proportional to  $M$  and converges to zero as  $N \rightarrow \infty$ . The value of the constant  $M$  depends on the choice for  $r_t$  and  $q_t$ . In particular, for the three forms of the EnKF algorithm, we have

$$M = \begin{cases} \Sigma_B \sup_t \mathbb{E}[\Sigma_t^{(N)}] + H^2 \sup_t \mathbb{E}[(\Sigma_t^{(N)})^3], & \text{(P-EnKF),} \\ \Sigma_B \sup_t \mathbb{E}[\Sigma_t^{(N)}], & \text{(S-EnKF),} \\ 0, & \text{(D-EnKF).} \end{cases}$$

The following conclusions can be drawn:

- (i) P-EnKF admits larger steady-state error compared to S-EnKF, while it was shown that it is more stable (see (12)).
- (ii) The steady-state error for D-EnKF, in approximating the variance, converges to zero, even for finite  $N$ . The reason is that the evolution of the empirical variance is deterministic, and identical to Kalman-filter equation, while for P-EnKF and S-EnKF involve stochastic terms.
- (iii) Assuming the conjecture (15) is true, the steady-state error for D-EnKF, in approximating the distribution, is proportional to  $d(\pi_0^{(N)}, \bar{\pi}_0)$  which depends on the initial conditions of the particles. If the initial particles are sampled randomly from  $\bar{\pi}_0$ , then the error decays as  $N^{-1}$ . However, if the initial particles are not random samples from  $\bar{\pi}_0$ , the error persists to exists even as  $N \rightarrow \infty$ . This is due to the fact that the mean-field system is not stable, as shown in (12). This is in contrast to P-EnKF and S-EnKF where the steady-state error for the distribution is proportional to  $\frac{M}{N}$  independent of the initial condition of the particles.

*Remark 5.* The result of the Prop. 3 also holds in vector-case, under additional assumption that the linear system is stable and fully observable ( $H$  is full-rank). For detailed analysis of P-EnKF algorithm, see [Del Moral and Tugaut 2018, Bishop and Del Moral 2018], and extension to nonlinear setting [Del Moral et al. 2017, de Wiljes et al. 2018]. Analysis of S-EnKF and D-EnKF appears in [Taghvaei and Mehta 2018b] and [Taghvaei and Mehta 2018a] respectively.

## 5. CONCLUSION

The paper presents stability and optimality analysis of EnKF algorithms. The central concept is the construction of a mean-field process that is exact, i.e. its time marginal distribution is equal to the filter posterior distribution. Because the exactness does not specify the joint distribution of the marginal distributions, there are infinity many exact mean-field processes (see Prop. 1). Stability and accuracy of three forms of mean-field process, that correspond to three forms of EnKF, are studied (see Table 1). It is shown that the deterministic EnKF is not stable, while stochastic forms of the EnKF are stable. The stability is stronger for EnKF with larger stochastic input (see Prop. 2 and the preceding discussion). While stochastic forms of EnKF are stable, they admit larger approximation error because of the presence of stochastic terms. The deterministic EnKF is accurate only if the particles are initialized properly,

because it does not correct for the initial error (see Prop. 3 and preceding discussion).

The analysis that is presented in this paper raises the question about how to design a mean-field process in the nonlinear and non-Gaussian setting, such that the mean-field process is stable.

## Appendix A. PROOF OF PROPOSITION 2

(i) To obtain a bound for the error in variance, we use the explicit solution to the Riccati equation (2b) that is available for the scalar case. In particular, let  $\phi_t(x)$  denote the semigroup associated with the Riccati equation such that  $\Sigma_t = \phi_t(\Sigma_0)$ . The explicit form of the semigroup is given by

$$\phi_t(x) = \frac{(\lambda_0 + A \tanh(\lambda_0 t))x + \Sigma_B \tanh(\lambda_0 t)}{\lambda_0 - A \tanh(\lambda_0 t) + H^2 \tanh(\lambda_0 t)x} \quad (\text{A.1})$$

The bound is obtained by providing a bound for the derivative of the semigroup, with respect to  $x$ , denoted by  $\nabla \phi_t(x)$ :

$$\begin{aligned} \nabla \phi_t(x) &= \frac{\lambda_0^2}{\cosh(\lambda_0 t)^2 (\lambda_0 - A \tanh(\lambda_0 t) + H^2 \tanh(\lambda_0 t)x)^2} \\ &\leq \frac{4\lambda_0^2}{(\lambda_0 - A)^2} e^{-2\lambda_0 t} \end{aligned} \quad (\text{A.2})$$

Therefore,  $\phi_t$  is globally Lipschitz and

$$|\bar{\Sigma}_t - \Sigma_t| = |\phi_t(\bar{\Sigma}_0) - \phi_t(\Sigma_0)| \leq c_1 e^{-2\lambda_0 t} |\bar{\Sigma}_0 - \Sigma_0|$$

where  $c_1 = \frac{4\lambda_0^2}{(\lambda_0 - A)^2}$ .

To prove the bound for the mean, we subtract the equation for  $m_t$  from the equation for  $\bar{m}_t$  to obtain

$$\begin{aligned} d\bar{m}_t - dm_t &= (A - \bar{\Sigma}_t H^2)(\bar{m}_t - m_t) dt \\ &\quad + (\bar{\Sigma}_t - \Sigma_t) H dI_t \end{aligned}$$

where  $dI_t := dZ_t - H\bar{m}_t dt$  is the innovation process. Therefore, the difference  $\bar{m}_t - m_t$  solves a linear system for which the solution is given by

$$\bar{m}_t - m_t = \psi_t(\bar{m}_0 - m_0) + \int_0^t \psi_{t,s}(\bar{\Sigma}_s - \Sigma_s) H dI_s$$

where  $\psi_{t,s} = \exp(\int_s^t (A - \bar{\Sigma}_\tau H^2) d\tau)$ . Next we obtain a bound for  $\psi_{t,s}$ .

$$\begin{aligned} \psi_{t,s} &= \exp\left(\int_s^t (A - \Sigma_\infty H^2) d\tau + \int_s^t (\Sigma_\infty - \Sigma_\tau) H^2 d\tau\right) \\ &\leq \exp(-(t-s)\lambda + c_1 |\bar{\Sigma}_0 - \Sigma_\infty| \int_s^t e^{-2\lambda\tau} d\tau) \\ &\leq c_2 \exp(-(t-s)\lambda) \end{aligned}$$

where  $c_2 = e^{\frac{c_1 |\Sigma_0 - \Sigma_\infty|}{2\lambda}}$ . Therefore,

$$\begin{aligned} \mathbb{E}[|\bar{m}_t - m_t|^2] &\leq c_2^2 e^{-2\lambda t} (\bar{m}_0 - m_0)^2 \\ &\quad + \int_0^t c_2^2 e^{-2\lambda(t-s)} |\bar{\Sigma}_s - \Sigma_s|^2 H^2 dt \\ &\leq c_2^2 e^{-2\lambda t} (\bar{m}_0 - m_0)^2 + c_3^2 e^{-2\lambda t} |\bar{\Sigma}_0 - \Sigma_0|^2 \end{aligned}$$

where  $c_3 = \frac{c_2 c_1 H}{\sqrt{2\lambda}}$ . Combining the bounds for the variance and the mean, we obtain the bound for the second term.

(ii) The result follows from a coupling argument. Introduce a new process  $\tilde{X}_t$  with the initial distribution  $N(\bar{m}_0, \bar{\Sigma}_0)$  governed by the same equation as the mean-field process:

$$\begin{aligned} d\tilde{X}_t &= \mathcal{T}_t(\bar{m}_t, \bar{\Sigma}_t) + G_t(\tilde{X}_t - \bar{m}_t) dt + r_t d\bar{B}_t \\ &\quad + q_t d\bar{W}_t, \quad \bar{X}_0 \sim N(\bar{m}_0, \bar{\Sigma}_0) \end{aligned}$$

where  $\bar{B}_t$  and  $\bar{W}_t$  are the same as in (8). It is straightforward to verify  $\bar{X}_t \sim \bar{\pi}_t = N(\bar{m}_t, \bar{\Sigma}_t)$ . Subtracting the equation for  $\bar{X}_t$  from  $\tilde{X}_t$  yields  $d(\bar{X}_t - \tilde{X}_t) = G_t(\bar{X}_t - \tilde{X}_t) dt$  concluding

$$\bar{X}_t - \tilde{X}_t = e^{\int_0^t G_s ds} (\bar{X}_0 - \tilde{X}_0)$$

The result follows from definition of the Wasserstein distance and the optimal coupling of the initial conditions.

(iii) The result follows from triangle inequality:

$$\mathcal{W}_2(\bar{\pi}_t, \pi_t) \leq \mathcal{W}_2(\bar{\pi}_t, \tilde{\pi}_t) + \mathcal{W}_2(\tilde{\pi}_t, \pi_t)$$

and application of part (i), part (ii), and of the formula for Wasserstein distance between two Gaussian distributions. ,

$$\mathcal{W}_2(\tilde{\pi}_t, \pi_t)^2 = (\bar{m}_t - m_t)^2 + (\sqrt{\bar{\Sigma}_t} - \sqrt{\Sigma_t})^2.$$

## Appendix B. PROOF OF THE PROPOSITION 3

The time evolution of the empirical variance  $\Sigma_t^{(N)}$  is given by:

$$d\Sigma_t^{(N)} = \text{Ricc}(\Sigma_t^{(N)}) dt + d\zeta_t^{(N)} \quad (\text{B.1})$$

where

$$d\zeta_t^{(N)} = \frac{2}{N-1} \sum_{i=1}^N (X_t^i - m_t^{(N)})(r_t dB_t^i + q_t dW_t^i).$$

Note that  $\zeta_t^{(N)}$  is a martingale and  $\langle d\zeta_t^{(N)} \rangle^2 = \frac{4(r_t^2 + q_t^2)}{N-1} \Sigma_t^{(N)} dt$ .

The difference  $\Sigma_t^{(N)} - \Sigma_t$  can be expressed, in terms of the semigroup (A.1), according to

$$\Sigma_t^{(N)} - \Sigma_t = \phi_0(\Sigma_t^{(N)}) - \phi_t(\Sigma_0^{(N)}) + \phi_t(\Sigma_0^{(N)}) - \phi_t(\Sigma_0)$$

The bound for the second term is straightforward:

$$|\phi_t(\Sigma_0^{(N)}) - \phi_t(\Sigma_0)| \leq c_1 e^{-2\lambda_0 t} |\Sigma_0^{(N)} - \Sigma_0|.$$

The first term is

$$\begin{aligned} \phi_0(\Sigma_t^{(N)}) - \phi_t(\Sigma_0^{(N)}) &= \int_0^t d_s \phi_{t-s}(\Sigma_s^{(N)}) \\ &= \int_0^t \nabla \phi_{t-s}(\Sigma_s^{(N)}) d\zeta_s^{(N)} + \int_0^t \frac{1}{2} \nabla^2 \phi_{t-s}(\Sigma_s^{(N)}) \langle d\zeta_s^{(N)} \rangle^2 \end{aligned}$$

where  $\nabla^2 \phi_t(x)$  denotes the second-order derivative with respect to  $x$ , and we used  $\frac{d}{dt} \phi_t(x) = \nabla \phi_t(x) \text{Ricc}(x)$  and (B.1). Therefore, using the bounds (A.2) and

$$|\nabla^2 \phi_t(x)| \leq c_4 e^{-2\lambda_0 t}$$

with  $c_4 = \frac{8\lambda_0^2 H^2}{(\lambda_0 - A)^3}$ , yields

$$\begin{aligned} \mathbb{E}[|\phi_0(\Sigma_t^{(N)}) - \phi_t(\Sigma_0^{(N)})|^2] &= \int_0^t \mathbb{E}[\nabla \phi_{t-s}(\Sigma_s^{(N)})^2 \langle d\zeta_s^{(N)} \rangle^2] \\ &+ \left[ \int_0^t \mathbb{E}[\nabla^2 \phi_{t-s}(\Sigma_s^{(N)}) \langle d\zeta_s^{(N)} \rangle^2] \right]^2 \\ &\leq \frac{4c_1^2}{N-1} \int_0^t e^{-4\lambda_0(t-s)} \mathbb{E}[(q_s^2 + r_s^2) \Sigma_s^{(N)}] ds \\ &+ \left[ \frac{4c_4}{N-1} \int_0^t e^{-2\lambda_0(t-s)} \mathbb{E}[(q_s^2 + r_s^2) \Sigma_s^{(N)}] ds \right]^2 \\ &\leq \frac{2c_1^2 M}{\lambda_0 N} \end{aligned}$$

if  $N > \frac{4c_1^2 M}{c_1^2 \lambda_0} = \frac{16H^4 M}{(\lambda_0 - A)^2 \lambda_0}$  and  $M := \sup_t \mathbb{E}[(q_t^2 + r_t^2) \Sigma_t^{(N)}]$ .

This concludes the bound:

$$\mathbb{E}[|\Sigma_t^{(N)} - \Sigma_t|^2] \leq \frac{4c_1^2 M}{\lambda_0 N} + 2c_1^2 e^{-2\lambda_0 t} \mathbb{E}[|\Sigma_0^{(N)} - \Sigma_0|^2]$$

- Abedi, E. and Surace, S.C. (2019). Gauge freedom within the class of linear feedback particle filters. *arXiv preprint arXiv:1903.06689*.
- Bergemann, K. and Reich, S. (2012). An ensemble Kalman-Bucy filter for continuous data assimilation. *Meteorologische Zeitschrift*, 21(3), 213–219. doi:10.1127/0941-2948/2012/0307.
- Bishop, A.N. and Del Moral, P. (2018). On the stability of matrix-valued Riccati diffusions. *arXiv preprint arXiv:1808.00235*. URL <https://arxiv.org/abs/1808.00235>.
- Chen, Y., Georgiou, T., and Pavon, M. (2016). Optimal steering of a linear stochastic system to a final probability distribution, part I. *IEEE Trans. Autom. Control*, 61(5), 1158–1169. doi:10.1109/TAC.2015.2457784.
- de Wiljes, J., Reich, S., and Stannat, W. (2018). Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise. *SIAM Journal on Applied Dynamical Systems*, 17(2), 1152–1181. doi:10.1137/17m1119056.
- Del Moral, P., Kurtzmann, A., and Tugaut, J. (2017). On the stability and the uniform propagation of chaos of a class of extended ensemble Kalman–Bucy filters. *SIAM Journal on Control and Optimization*, 55(1), 119–155. doi:10.1137/16M1087497.
- Del Moral, P. and Tugaut, J. (2018). On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. *Ann. Appl. Probab.*, 28(2), 790–850. doi:10.1214/17-AAP1317. URL <https://doi.org/10.1214/17-AAP1317>.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5), 10143–10162. doi:10.1029/94JC00572.
- Evensen, G. (2006). *Data Assimilation. The Ensemble Kalman Filter*. Springer-Verlag, New York.
- Houtekamer, P. and Mitchell, H. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, 129, 123–136.
- Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction theory. *Journal of basic engineering*, 83(1), 95–108. doi:10.1115/1.3658902.
- Kelly, D., Law, K.J., and Stuart, A.M. (2014). Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27(10), 2579. doi:10.1088/0951-7715/27/10/2579.
- Kwiatkowski, E. and Mandel, J. (2015). Convergence of the square root ensemble Kalman filter in the large ensemble limit. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1), 1–17.
- Le Gland, F., Monbet, V., and Tran, V. (2009). *Large sample asymptotics for the ensemble Kalman filter*. Ph.D. thesis, INRIA.
- Mandel, J., Cobb, L., and Beezley, J.D. (2011). On the convergence of the ensemble Kalman filter. *Applications of Mathematics*, 56(6), 533–541.
- Ocone, D. and Pardoux, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM Journal on Control and Optimization*, 34(1), 226–243. doi:10.1137/s0363012993256617.
- Reich, S. (2011). A dynamical systems framework for intermittent data assimilation. *BIT Numerical Analysis*, 51, 235–249.

doi:10.1007/s10543-010-0302-4.

- Taghvaei, A. and Mehta, P.G. (2016). An optimal transport formulation of the linear feedback particle filter. In *American Control Conference (ACC), 2016*, 3614–3619. IEEE. doi: 10.1109/acc.2016.7525474.
- Taghvaei, A. (2019). *Design and analysis of particle-based algorithms for nonlinear filtering and sampling*. Ph.D. thesis, University of Illinois at Urbana-Champaign.
- Taghvaei, A. and Mehta, P.G. (2018a). Error analysis for the linear feedback particle filter. In *2018 Annual American Control Conference (ACC)*, 4261–4266. IEEE. doi:10.23919/ACC.2018.8430867.
- Taghvaei, A. and Mehta, P.G. (2018b). Error analysis of the stochastic linear feedback particle filter. In *2018 IEEE Conference on Decision and Control (CDC)*, 7194–7199. IEEE.
- Taghvaei, A. and Mehta, P.G. (2020). An optimal transport formulation of the ensemble kalman filter. *IEEE Transactions on Automatic Control*.
- Tong, X.T., Majda, A.J., and Kelly, D. (2016). Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity*, 29(2), 657.
- Villani, C. (2003). *Topics in Optimal Transportation*, volume 58. American Mathematical Soc. doi:10.1090/gsm/058/09.
- Whitaker, J. and Hamill, T.M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7), 1913–1924. doi:10.1175/1520-0493(2002)130<1913:edawpo>2.0.co;2.
- Yang, T., Laugesen, R.S., Mehta, P.G., and Meyn, S.P. (2016). Multivariable feedback particle filter. *Automatica*, 71, 10–23. doi:10.1016/j.automatica.2016.04.019.
- Yang, T., Mehta, P.G., and Meyn, S.P. (2013). Feedback particle filter. *IEEE Transactions on Automatic Control*, 58(10), 2465–2480. doi:10.1109/TAC.2013.2258825.