Maximum Correntropy Ensemble Kalman Filter

Yangtianze Tao, Jiayi Kang and Stephen Shing-Tong Yau

Abstract— In this article, a robust ensemble Kalman filter (EnKF) called MC-EnKF is proposed for nonlinear state-space model to deal with filtering problems with non-Gaussian observation noises. Our MC-EnKF is derived based on maximum correntropy criterion (MCC) with some technical approximations. Moreover, we propose an effective adaptive strategy for kernel bandwidth selection. Besides, the relations between the common EnKF and MC-EnKF are given, i.e., MC-EnKF will converge to the common EnKF when the kernel bandwidth tends to infinity. This justification provides a complementary understanding of the kernel bandwidth selection for MC-EnKF. In experiments, non-Gaussian observation noises significantly reduce the performance of the common EnKF for both linear and nonlinear systems, whereas our proposed MC-EnKF with a suitable kernel bandwidth maintains its good performance at only a marginal increase in computing cost, demonstrating its robustness and efficiency to non-Gaussian observation noises.

I. INTRODUCTION

Filtering [1] is the fundamental issue for the state estimation of the state-space model [2]. It has a large number of applications in many fields, such as robot vision [3] and data assimilation [4]. For the linear state-space model with Gaussian noise, the optimal filtering solution can be computed analytically, which is the well-known Kalman filter (KF) [5]. Since then, a large number of filtering algorithms for nonlinear state-space model with Gaussian noise have been proposed. The representatives among them are extended Kalman filter (EKF) [6], unscented Kalman filter (UKF) [7], cubature Kalman filter (CKF) [8] and ensemble Kalman filter (EnKF) [9].

In the case of Gaussian noise, the KF and its expansions work well. However, the noise frequently does not fit the Gaussian distribution in real-world application circumstances. For instance, in many practical settings of target tracking [10] and power systems [11], impulsive interferences and observation outliers are frequent. These disturbances are often modelled by some heavy-tailed impulsive noises (such as some mixed-Gaussian distributions). The main reason for this problem is that the KF is based on the well-known minimum mean square error (MMSE) criterion [12], which is sensitive to large outliers and deteriorates

the KF's robustness in non-Gaussian noise [13]. Numerous studies have attempted to address the filtering problems with non-Gaussian noise, such as filters based on information theoretic quantities [14], the Huber-based KFs [15] and robust Student's t filter [16]. Besides these filters, a local similarity measure called correntropy [17] has been used to develop new robust filters [18], [19], [20], [21]. These filters are derived by maximum correntropy criterion (MCC) and called MCC-based filters, which can achieve great performance in the presence of heavy-tailed non-Gaussian noises since correntropy is insensitive to large outliers.

In this article, we formulate the estimation problem under MCC framework. Under this framework, the goal of this article is to develop a robust EnKF called the maximum correntropy EnKF (MC-EnKF) based on the MCC, to address nonlinear filtering problems when observations contain large outliers. With the help of suitable approximation techniques, we derive the recursion of ensembles for MC-EnKF based on MCC cost function. It is necessary to note that MC-EnKF employs the empirical prediction covariance calculated by ensembles. This procedure is consistent with the common EnKF, which allows us to prove that MC-EnKF will converge to the common EnKF when the kernel bandwidth tends to infinity. This justification provides a complementary knowledge of understanding the kernel bandwidth selection for MC-EnKF. Although MC-EnKF aims to handle non-Gaussian noise cases, it also has the flexibility to handle Gaussian noise cases with a large enough kernel bandwidth since EnKF performs well in this scenario. The contributions of this article are listed as follows.

- We introduce the idea of the cost function to the field of EnKF, which inspires us to derive a robust EnKF called MC-EnKF based on the MCC cost function. Moreover, we propose an effective adaptive strategy for kernel bandwidth selection. Besides, we provide a complementary understanding of the kernel bandwidth selection theoretically, i.e., MC-EnKF will converge to the common EnKF when the kernel bandwidth is large enough. This justification gives the flexibility of MC-EnKF to handle cases involving both Gaussian and non-Gaussian noise.

- Our theoretical result is supported by numerical experiments on several filtering benchmarks, i.e., our proposed MC-EnKF will perform like the common EnKF when the kernel bandwidth is large enough. Furthermore, experiments demonstrate that our proposed adaptive strategy for kernel bandwidth selection is effective. And with the appropriate kernel bandwidth, MC-EnKF

can outperform EnKF with only a minor increase in computing cost when the underlying observation system is hampered by some heavy-tailed non-Gaussian noises.

The remainder of this article is organized as follows. In section II, we present the concept of MCC, and briefly review the nonlinear filtering problems and EnKF algorithm. In section III, we derive our MC-EnKF algorithm and propose an effective adaptive strategy for kernel bandwidth selection with necessary discussions. The experiments and discussions are presented in Section IV. The conclusions are drawn in Section V.

A. Notations

Throughout the article, we use boldface lower-case letters for vectors and boldface upper-case letters for matrices. The transpose and mathematical expectation are denoted by $\{\cdot\}^\top$ and $\mathbb{E}[\cdot]$, respectively. The Gaussian distribution with mean μ and covariance $\mathbf{\Sigma}$ is denoted by $\mathcal{N}(\mu, \mathbf{\Sigma})$. $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ denote, respectively, the n-dimensional Euclidean space and the set of all $n \times m$ real matrices. Especially, $\mathbf{I}_n$ denotes the n-dimensional identity matrix. weighted l^2 norm $\|\mathbf{x}\|_{\mathbf{A}}^2 := \mathbf{x}^\top \mathbf{A}^{-1} \mathbf{x}$ with $\mathbf{x} \in \mathbb{R}^n$ and non-singular matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$.

II. BACKGROUND

In this section, we shall first introduce the concept of correntropy [17], which is a local similarity measure between two random vectors. Then we shall formulate the estimation problems based on MCC and MMSE estimate [12]. At last, we shall briefly review the framework of nonlinear filtering problems and introduce the EnKF algorithm.

A. Maximum Correntropy Criterion and Minimum Mear Square Error

1) Correntropy: For given two random vectors $\mathbf{x}$ and $\mathbf{y}$, the correntropy between $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathcal{V}(\mathbf{x}, \mathbf{y})$, which is defined as follows:

$$\mathcal{V}(\mathbf{x}, \mathbf{y}) = \mathbb{E}\left[\mathcal{G}_\sigma(\|\mathbf{x} - \mathbf{y}\|_{\mathbf{A}_1})\right], \quad (1)$$

where $\mathbf{A}_1$ is a given non-singular matrix with suitable dimension, $\mathcal{G}_\sigma$ is the Gaussian Kernel given by $\mathcal{G}_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right)$, and $\sigma > 0$ stands for the kernel bandwidth. Here we shall recall the important property for $\mathcal{V}(\mathbf{x}, \mathbf{y})$, more details can be found in [17]. By taking the Taylor series expansion of the Gaussian Kernel, we have

$$\mathcal{V}(\mathbf{x}, \mathbf{y}) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} \mathbb{E}\left[\|\mathbf{x} - \mathbf{y}\|_{\mathbf{A}_1}^{2n}\right]. \quad (2)$$

The correntropy can be seen as the weighted sum of all even order moments of the error vectors $\mathbf{x} - \mathbf{y}$. The parameter to weight the second order and higher order moments appears to be the kernel bandwidth. The second order moment will predominate in the correntropy when the kernel bandwidth is particularly large compared to the error vectors range.

2) Estimation Problems Formulation: Correntropy can be used as the optimality criterion for estimation problems. Suppose our goal is to learn a parameter θ for a given estimator $\mathbf{x}(\theta)$, and let $\mathbf{y}$ denote the desired output. Then the MCC-based estimation problem can be formulated as solving the following optimzation problem:

$$\hat{\boldsymbol{\theta}}_1 = \arg \max_{\boldsymbol{\theta} \in \Omega_1} \mathcal{V}(\mathbf{x}(\theta), \mathbf{y}), \quad (3)$$

where $\hat{\boldsymbol{\theta}}_1$ denotes its optimal solution and Ω_1 denotes its feasible set for parameters. It is necessary to compare MCC with the conventional MMSE criterion. Let Ω_2 denotes its feasible set for parameters with $\hat{\boldsymbol{\theta}}_2$ as optimal solution. Then for the given non-singular matrix $\mathbf{A}_2$ with suitable dimension, the MMSE-based estimation problem is formulated as follows:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_2 &= \arg \max_{\boldsymbol{\theta} \in \Omega_2} -\|\mathbf{x}(\theta) - \mathbf{y}\|_{\mathbf{A}_2} \\ &= \arg \min_{\boldsymbol{\theta} \in \Omega_2} \|\mathbf{x}(\theta) - \mathbf{y}\|_{\mathbf{A}_2}. \end{aligned} \quad (4)$$

B. Nonlinear Filtering Problems

In this article, we consider the nonlinear state-space model given by the following state and observation equations:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \end{aligned} \quad (5)$$

where state $\mathbf{x}_k \in \mathbb{R}^n$, and observation $\mathbf{y}_k \in \mathbb{R}^m$. $\mathbf{f}_k$ and $\mathbf{h}_k$ are nonlinear functions called state equation and observation equation, respectively. State noise $\mathbf{w}_k$ and observation noise $\mathbf{v}_k$ are zero means with nominal covariance $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$ and $\mathbf{R}_k \in \mathbb{R}^{m \times m}$, respectively. Let $\mathbf{y}_{1:k}$ denote the σ-algebra generated by noisy observations $\{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$. The filtering problem refers to estimating the posterior distribution $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$, which is called filtering distribution.

In the follow-up, we focus on the scenario in which the observation noises are not Gaussian, i.e., they are induced by unknown large outliers. Hence, the real distribution of observation noise is unknown to us in fact. For example, we consider $(1 - \epsilon)\mathcal{N}(0, \mathbf{R}_k) + \epsilon \mathcal{S}(0, \mathbf{S}_k)$, where $0 < \epsilon \ll 1$ is the unknown probability, and $\mathcal{S}(0, \mathbf{S}_k)$ is an arbitrary unknown distribution with large covariance $\mathbf{S}_k$. Here $\mathbf{R}_k$ is known to us, so it is called the nominal covariance matrix. Additionally, in what follows, we shall develop our novel filter to deal with such noises above by utilizing the MCC as the cost function involving the nominal observation covariance $\mathbf{R}_k$.

C. Ensemble Kalman Filter

The EnKF sequentially approximates the filtering distributions $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$ using N equally weighted ensembles $\{\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(N)}\}$. At prediction steps, each ensemble $\mathbf{x}_k^{(i)}$ is propagated using the state equation, while at update steps a Kalman-type update is performed for each ensemble:

$$\mathbf{x}_{k|k-1}^{(i)} = \mathbf{f}_k(\mathbf{x}_{k-1|k-1}^{(i)}) + \mathbf{w}_k^{(i)}, \quad (6)$$

and

$$\mathbf{x}_{k|k}^{(i)} = \mathbf{x}_{k|k-1}^{(i)} + \hat{\mathbf{K}}_k \left(\mathbf{y}_k + \mathbf{v}_k^{(i)} - \mathbf{h}_k(\mathbf{x}_{k|k-1}^{(i)})\right), \quad (7)$$

where $\mathbf{w}_k^{(i)} \sim \mathcal{N}(0, \mathbf{Q}_k)$, $\mathbf{v}_k^{(i)} \sim \mathcal{N}(0, \mathbf{R}_k)$ and the Kalman gain

$$\hat{\mathbf{K}}_k = \hat{\mathbf{C}}_k \mathbf{H}_k^\top (\mathbf{H}_k \hat{\mathbf{C}}_k \mathbf{H}_k^\top + \mathbf{R}_k)^{-1}, \tag{8}$$

is defined using the empirical prediction covariance $\hat{\mathbf{C}}_k$ of prediction ensembles $\{\mathbf{x}_{k|k-1}^{(1)}, \dots, \mathbf{x}_{k|k-1}^{(N)}\}$, namely

$$\hat{\mathbf{C}}_k = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_{k|k-1}^{(i)} - \hat{\mathbf{m}}_k)(\mathbf{x}_{k|k-1}^{(i)} - \hat{\mathbf{m}}_k)^\top, \tag{9}$$

with

$$\hat{\mathbf{m}}_k = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{k|k-1}^{(i)}, \tag{10}$$

and

$$\mathbf{H}_k = \frac{\partial \mathbf{h}_k}{\partial \mathbf{x}_k} \mid_{\mathbf{x}_k = \hat{\mathbf{m}}_k}. \tag{11}$$

Remark 2.1: Empirical prediction mean in (10) and empirical prediction covariance in (9) provide a Gaussian approximation to the prediction distribution distribution, i.e.,

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) \approx \mathcal{N}(\hat{\mathbf{m}}_k, \hat{\mathbf{C}}_k). \tag{12}$$

III. PROPOSED ALGORITHM

In this section, we shall present the derivation of MC-EnKF. Then we shall give the discussions on the adaptive strategy for kernel bandwidth selection and the convergence of MC-EnKF with respect to kernel bandwidth.

A. Derivation of the Algorithm

Here we present how to derive MC-EnKF based on the MCC estimate (3). The prediction step of MC-EnKF is the same as those of EnKF, i.e., (6). Therefore, we shall focus on deriving its update step. Let $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ denote the prediction mean $\mathbb{E}[\mathbf{x}_k \mid \mathbf{y}_{1:k-1}]$ and prediction covariance $\mathbb{E}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^\top \mid \mathbf{y}_{1:k-1}\right]$, respectively. As shown in [22], for the underlying linear system of (5), i.e., $\mathbf{f}_k(\mathbf{x}_k) = \mathbf{F}_k \mathbf{x}_k$ and $\mathbf{h}_k(\mathbf{x}_k) = \mathbf{H}_k \mathbf{x}_k$, the update step of KF can be derived by using least square cost function,

$$\mathcal{L}_1(\mathbf{x}_k) = \|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|_{\mathbf{R}_k} + \|\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\|_{\mathbf{P}_{k|k-1}}. \tag{13}$$

Motivated by (13), we shall consider a new cost function based on MCC. Besides, in view of Remark 2.1, we shall replace $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ in (13) by the empirical prediction mean $\hat{\mathbf{m}}_k$ in (10) and the empirical prediction covariance $\hat{\mathbf{C}}_k$ in (9). Therefore, our modified cost function for ensembles $\mathbf{x}_{k|k}^{(i)}$ for $i = 1, 2, \dots, N$ is given by

$$\mathcal{L}_2(\mathbf{x}_k) = \mathcal{G}_\sigma \left(\|\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)\|_{\mathbf{R}_k}\right) + \mathcal{G}_\sigma \left(\|\mathbf{x}_k - \hat{\mathbf{m}}_k\|_{\hat{\mathbf{C}}_k}\right). \tag{14}$$

Then based on $\mathcal{L}_2(\mathbf{x}_k)$, the update step for MC-EnKF can be obtained solving the following optimization problem for $i = 1, 2, \dots, N$:

$$\mathbf{x}_{k|k}^{(i)} = \arg \max_{\mathbf{x}_k} \ \mathcal{L}_2(\mathbf{x}_k). \tag{15}$$

In what follows, the derivation contains some approximations. For the sake of obtaining our algorithm, we heuristically treat them as exact equalities. Let us denote

$$\begin{aligned} l_{\mathbf{R}_k} &= \mathcal{G}_\sigma \left(\|\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)\|_{\mathbf{R}_k}\right) \\ l_{\hat{\mathbf{C}}_k} &= \mathcal{G}_\sigma \left(\|\mathbf{x}_k - \hat{\mathbf{m}}_k\|_{\hat{\mathbf{C}}_k}\right). \end{aligned} \tag{16}$$

Then recall $\mathbf{H}_k$ defined in (11), we consider this approximation when taking the derivative,

$$\begin{aligned} \frac{\partial \mathcal{L}_2(\mathbf{x}_k)}{\partial \mathbf{x}_k} &= -\frac{1}{\sigma^2} \left(\frac{\partial \mathbf{h}_k}{\partial \mathbf{x}_k}\right)^\top l_{\mathbf{R}_k} \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)) \\ &\quad + \frac{1}{\sigma^2} l_{\hat{\mathbf{C}}_k} \hat{\mathbf{C}}_k^{-1} (\mathbf{x}_k - \hat{\mathbf{m}}_k) \\ &\approx -\frac{1}{\sigma^2} \mathbf{H}_k^\top l_{\mathbf{R}_k} \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)) \\ &\quad + \frac{1}{\sigma^2} l_{\hat{\mathbf{C}}_k} \hat{\mathbf{C}}_k^{-1} (\mathbf{x}_k - \hat{\mathbf{m}}_k). \end{aligned} \tag{17}$$

Now letting $\frac{\partial \mathcal{L}_2(\mathbf{x}_k)}{\partial \mathbf{x}_k} = 0$, we have

$$\mathbf{H}_k^\top l_{\mathbf{R}_k} \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)) = l_{\hat{\mathbf{C}}_k} \hat{\mathbf{C}}_k^{-1} (\mathbf{x}_k - \hat{\mathbf{m}}_k). \tag{18}$$

Adopting the first-order Taylor series to approximate the nonlinear observation function $\mathbf{h}_k$ at $\mathbf{x}_{k|k-1}^{(i)}$, i.e.,

$$\mathbf{h}_k(\mathbf{x}_k) \approx \mathbf{h}_k\left(\mathbf{x}_{k|k-1}^{(i)}\right) + \mathbf{H}_k\left(\mathbf{x}_k - \mathbf{x}_{k|k-1}^{(i)}\right). \tag{19}$$

Substituting (19) into (18), we have

$$\begin{aligned} \mathbf{x}_k = \mathbf{x}_{k|k-1}^{(i)} &+ \left(l_{\hat{\mathbf{C}}_k} \hat{\mathbf{C}}_k^{-1} + \mathbf{H}_k^\top l_{\mathbf{R}_k} \mathbf{R}_k^{-1} \mathbf{H}_k\right)^{-1} \\ &\times \mathbf{H}_k^\top l_{\mathbf{R}_k} \mathbf{R}_k^{-1} \left(\mathbf{y}_k - \mathbf{h}_k\left(\mathbf{x}_{k|k-1}^{(i)}\right)\right). \end{aligned} \tag{20}$$

Then we obtain the following stochastic ensemble update rule like (7) for $\mathbf{x}_{k|k}^{(i)}$ with drawing $\mathbf{v}_k^{(i)} \sim \mathcal{N}(0, \mathbf{R}_k)$:

$$\mathbf{x}_{k|k}^{(i)} = \mathbf{x}_{k|k-1}^{(i)} + \tilde{\mathbf{K}}_k \left(\mathbf{y}_k + \mathbf{v}_k^{(i)} - \mathbf{h}_k\left(\mathbf{x}_{k|k-1}^{(i)}\right)\right), \tag{21}$$

where the new Kalman gain $\tilde{\mathbf{K}}_k$ is given by

$$\begin{aligned} \tilde{\mathbf{K}}_k &= \left(l_{\hat{\mathbf{C}}_k} \hat{\mathbf{C}}_k^{-1} + \mathbf{H}_k^\top l_{\mathbf{R}_k} \mathbf{R}_k^{-1} \mathbf{H}_k\right)^{-1} \mathbf{H}_k^\top l_{\mathbf{R}_k} \mathbf{R}_k^{-1} \\ &= \left(\left(\frac{\hat{\mathbf{C}}_k}{l_{\hat{\mathbf{C}}_k}}\right)^{-1} + \mathbf{H}_k^\top \left(\frac{\mathbf{R}_k}{l_{\mathbf{R}_k}}\right)^{-1} \mathbf{H}_k\right)^{-1} \mathbf{H}_k^\top \left(\frac{\mathbf{R}_k}{l_{\mathbf{R}_k}}\right)^{-1}. \end{aligned} \tag{22}$$

Here we shall discuss the practical algorithm, since $l_{\mathbf{R}_k}$ and $l_{\hat{\mathbf{C}}_k}$ all contain the item $\mathbf{x}_k$. Here we introduce a novel technical approximation. Recall (16), we use $\hat{\mathbf{m}}_k$ to approximate $\mathbf{x}_k$ contained in $l_{\mathbf{R}_k}$ and $l_{\hat{\mathbf{C}}_k}$, respectively, i.e.,

$$\begin{aligned} \hat{l}_{\mathbf{R}_k} &= \mathcal{G}_\sigma \left(\|\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{m}}_k)\|_{\mathbf{R}_k}\right) \approx l_{\mathbf{R}_k} \\ \hat{l}_{\hat{\mathbf{C}}_k} &= \mathcal{G}_\sigma \left(\|\hat{\mathbf{m}}_k - \hat{\mathbf{m}}_k\|_{\hat{\mathbf{C}}_k}\right) \approx l_{\hat{\mathbf{C}}_k}. \end{aligned} \tag{23}$$

It follows that we summarise the specific algorithm steps of MC-EnKF in Algorithm 1.

Remark 3.1: The proposed MC-EnKF is still of Kalman type, i.e., it has a similar recursive structure as the common EnKF. Moreover, its computational complexity is only

---

Algorithm 1 MC-EnKF

1: **Intitialization.** Start with initial filtering ensembles $\mathbf{x}_{0|0}^{(1)}, \ldots, \mathbf{x}_{0|0}^{(N)}$. Then, at each time $k = 1, 2, \ldots$, given filtering ensembles $\mathbf{x}_{k-1|k-1}^{(1)}, \ldots, \mathbf{x}_{k-1|k-1}^{(N)}$, the MC-EnKF carries out the following two steps for $i = 1, 2, \ldots, N$:

2: **Prediction Step:** Draw $\mathbf{w}_k^{(i)} \sim \mathcal{N}(0, \mathbf{Q}_k)$ and calculate $\mathbf{x}_{k|k-1}^{(i)}$ via (6).

3: **Update Step:** Draw $\mathbf{v}_k^{(i)} \sim \mathcal{N}(0, \mathbf{R}_k)$ and calculate $\mathbf{x}_{k|k}^{(i)}$ by

$$\mathbf{x}_{k|k}^{(i)} = \mathbf{x}_{k|k-1}^{(i)} + \tilde{\mathbf{K}}_k \left(\mathbf{y}_k + \mathbf{v}_k^{(i)} - \mathbf{h}_k \left(\mathbf{x}_{k|k-1}^{(i)} \right) \right),$$

where $\tilde{\mathbf{K}}_k$ is defined in (22) with the approximation (23).

---

slightly higher than the common EnKF, which is verified in later experiments.

Remark 3.2: In (22), we note that MC-EnKF is potential to handle non-Gaussian observation noises thanks to the two weights, $\hat{l}_{\hat{\mathbf{C}}_k}$ and $\hat{l}_{\mathbf{R}_k}$, which could adjust $\hat{\mathbf{C}}_k$ and $\mathbf{R}_k$ adaptively. We shall also verify this argument in later experiments.

B. Adaptive Strategy for Kernel Bandwidth Selection and Convergence Regarding to Kernel Bandwidth

In general, the kernel bandwidth (scale parameter) in the cost function balances the convergence rates of the regression model and its robustness [23]. Since our algorithm does not involve solving such regression problems, we focus on the how to tune this scale parameter for robustness (with respect to outliers). Intuitively, if the arrived observation $\mathbf{y}_k$ contains large outliers, $\|\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_k)\|_2$ will be large, where $\|\cdot\|_2$ denote the l^2-norm. In this case, we need a smaller σ to make our algorithm more robust. Motivated by this intuition, we propose to set $\sigma_k = \frac{1}{\|\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_k)\|_2}$ adaptively, where σ_k denote the kernel bandwidth used in the k-th step for MC-EnKF. As shown in our simulation experiments, this adaptive strategy can achieve a good estimation performance. Additionally, the MC-EnKF will act more and more like the common EnKF algorithm as σ increases. In particular, the following convergence theorem holds .

Theorem 3.1: If the kernel bandwidth $\sigma \to \infty$, the proposed MC-EnKF will converge to the common EnKF.

Proof: See Appendix I. ∎

Remark 3.3: This justification in Theorem 3.1 gives the flexibility of MC-EnKF to handle Gaussian noise cases since the common EnKF performs well in this case.

IV. EXPERIMENTS

In this section, we will conduct performance comparisons of our proposed MC-EnKF with the common EnKF and the maximum liklihood EnKF (ML-EnKF) [24], which optimizes a nonlinear cost function through maximum likelihood. These evaluations will be carried out on various filtering

benchmarks that incorporate non-Gaussian noises, i.e., the observation noises with large outliers. In these experiments, we consider $M = 100$ independent Monte Carlo runs. In each run, $N = 1000$ samples are used to evaluate the MSE of the state. The EnKF and MC-EnKF are implemented by Numpy [25] and run on Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz.

A. Linear System

The first benchmark we used here is a linear system, which is given by

$$\mathbf{x}_k = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{w}_k,$$
$$\mathbf{y}_k = \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k, \tag{24}$$

where $\alpha = \frac{\pi}{18}$, $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I}_2)$ and $\mathbf{w}_k \sim \mathcal{N}(0, 0.01\mathbf{I}_2)$. Let $\mathbf{Q}_1 = 0.01\mathbf{I}_1$ be the nominal observation covariance, the observation noises are sampled from the mixture of Gaussian, i.e.,

$$\mathbf{v}_k \sim 0.9 \, \mathcal{N}(0, \mathbf{Q}_1) + 0.1 \, \mathcal{N}(0, 100\mathbf{Q}_1). \tag{25}$$

Table I lists the MSEs and average CPU times of EnKF and MC-EnKF with different kernel bandwidths in this example, where the number of ensembles of EnKF and MC-EnKF are set as 100. Note that the MC-EnKF outperforms EnKF in most cases and their average CPU times are virtually identical to the common EnKF. This demonstrates that our proposed MC-EnKF has better performance than those of EnKF with nearly no extra cost on CPU time. It means that MC-EnKF is more robust and efficient when we deal with observation noises with large outliers. In fact, our proposed adaptive strategy (MC-EnKF-Ada) outperforms most different σ cases and is only slightly worse than the best case. This suggests that our adaptive strategy works well for linear system. When we conduct the simulation for large enough σ, for example, $\sigma = 10000$, the result verifies Theorem 3.1 for linear system since we note that the performance of MC-EnKF is almost identical to the EnKF in this case.

TABLE I

THE MSES COMPARISONS BETWEEN ENKF AND MC-ENKF WITH DIFFERENT σ FOR LINEAR SYSTEM.

Methods	MSE	CPU Time
EnKF	3.1004	0.4676
MC-EnKF-Ada	2.2344	0.4697
MC-EnKF ($\sigma = 0.1$)	2.9655	0.4634
MC-EnKF ($\sigma = 0.5$)	2.9247	0.4636
MC-EnKF ($\sigma = 2$)	2.1031	0.4689
MC-EnKF ($\sigma = 5$)	**1.9411**	0.4629
MC-EnKF ($\sigma = 10$)	2.3073	0.4642
MC-EnKF ($\sigma = 10000$)	3.1605	0.4725

In order to better present the comparison results, we choose the case where $\sigma = 5$ to illustrate the true state and EnKF estimation and MC-EnKF estimation over time, which is shown in Fig 1. Here, the selected dimension is indicated by the - suffix of the legend in the figures; for example, EnKF-1 denotes the EnKF estimation on the

first state dimension. This setting will also be used for the following figure. We can clearly see that this MC-EnKF gives accurate estimates but EnKF performs poorly when the observations contain large outliers.
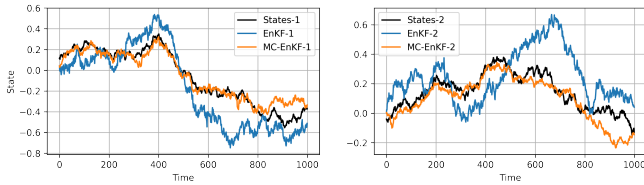

Fig. 1. The true state versus the estimate of EnKF and the estimate of MC-EnKF ($\sigma = 5$) over time for linear system.

B. Nonlinear System

The second benchmark is a nonlinear system, which is given as follows:

$$\mathbf{x}_k = \left(\mathbf{I}_2 + \kappa_1 \begin{bmatrix} -1 & 0.2 \\ 0.2 & -1 \end{bmatrix}\right)\mathbf{x}_{k-1} + \kappa_2 \cos(\mathbf{x}_{k-1}) + \mathbf{w}_k,$$
$$\mathbf{y}_k = \mathbf{x}_k + \sin(\mathbf{x}_k) + \mathbf{v}_k,$$
$$(26)$$

where $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I}_2)$ and $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{I}_2)$. $\kappa_1 = \kappa_2 = 0.1$ are constants controlling the state dynamics. Let $\mathbf{Q}_2 = \mathbf{I}_2$ be the nominal observation covariance, the observation noises are sampled from the mixture of Gaussian, i.e.,

$$\mathbf{v}_k \sim 0.9 \, \mathcal{N}(0, \mathbf{Q}_2) + 0.1 \, \mathcal{N}(0, 1000\mathbf{Q}_2). \qquad (27)$$

We list the MSEs and average CPU times of EnKF, ML-EnKF and MC-EnKF with different kernel bandwidths in Table II for this nonlinear example, where the number of ensembles of these filters still be set as 100. It is obvious that for nonlinear system the EnKF degrades greatly in the presence of non-Gaussian observation noises and the linear approximation error of observation function, hence has the worst estimate performance. ML-EnKF performs better than EnKF, but is still affected by outliers. Better results can be obtained with our proposed adaptive strategy (MC-EnKF-Ada) than with some kernel bandwidths, indicating that it is still useful for nonlinear systems. Note that the MC-EnKF outperforms EnKF in all cases, which demonstrate that our proposed MC-EnKF can effectively eliminate the effect of the non-Gaussian observation noises. Additionally, we notice that they share almost identical CPU times, which demonstrates the efficiency of MC-EnKF. Moreover, with suitable kernel bandwidth ($\sigma = 5$), MC-EnKF can achieve very accurate estimates while EnKF is heavily affected by large outliers, which is presented in Fig 2. We also note that when σ is large enough ($\sigma = 10000$), the performance of MC-EnKF is identical to the EnKF, which verifies the result of Theorem 3.1 for nonlinear system.

C. Discussions

According to the simulation experiments mentioned above, we notice the following two arguments:

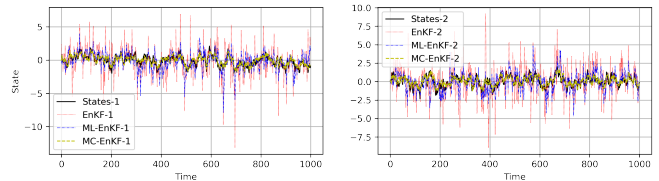| Methods	MSE	CPU Time
EnKF	4.0929	0.4768
ML-EnKF	3.3567	0.5059
MC-EnKF-Ada	2.9282	0.4775
MC-EnKF ($\sigma = 0.1$)	3.0150	0.4793
MC-EnKF ($\sigma = 0.5$)	2.9703	0.4769
MC-EnKF ($\sigma = 2$)	1.5849	0.4713
MC-EnKF ($\sigma = 5$)	**1.3012**	0.4762
MC-EnKF ($\sigma = 10$)	1.3752	0.4729
MC-EnKF ($\sigma = 10000$)	4.0448	0.4799

Fig. 2. The true state versus the estimate of EnKF and the estimate of MC-EnKF ($\sigma = 5$) over time for nonlinear system.

- With suitable kernel bandwidth, MC-EnKF can significantly improve its robustness to observation noises containing large outliers with nearly no CPU time loss.
- With large kernel bandwidth, MC-EnKF performs like the common EnKF, which implies its flexibility to handle Gaussian noise cases.

These two arguments make our viewpoint a very fruitful area for further study. Besides, we note that the kernel bandwidth σ plays a significant role in the algorithm and will have an impact on its robustness to non-Gaussian observation (such as outliers). Although our proposed adaptive strategy is effective, it cannot achieve the performance with suitable kernel bandwidth. Therefore, a highly intriguing future work direction may be a more effective adaptive strategy to pick the right kernel bandwidth with theoretical guarantee since it is crucial for our proposed MC-EnKF. In Addition, this article only investigates using the special case (MCC) of generalized MCC [26] as a cost function to derive a robust EnKF with stochastic update steps. Hence how to develop robust variants of EnKF [9] based on the generalized MCC will be our next research focus.

V. CONCLUSION

This article proposes a robust EnKF filtering algorithm called maximal correntropy ensemble Kalman filter (MC-EnKF), which is flexible to handle cases involving both Gaussian and non-Gaussian noise. Instead of the well-known minimum mean square error (MMSE) criterion, the MC-EnKF is derived by employing the maximum correntropy criterion (MCC) as the optimality criterion. Ensemble propagation equations remain to be of the Kalman type. The MC-EnKF will behave like the EnKF when the kernel bandwidth is large enough, and we also theoretically demonstrate this

argument. With the proper kernel bandwidth, MC-EnKF can perform much better than the EnKF at only a slight increase in computational cost, especially when the underlying observation system is disturbed by some heavy-tailed non-Gaussian noises. Besides, we propose an adaptive strategy to help us choose kernel bandwidth, and its effectiveness is also been verified by experiments.

APPENDIX I
PROOF OF THE THEOREM 3.1

Here we shall give the proof of the Theorem 3.1. Before proceeding, we need the following technical lemma.

Lemma 1.1 (Matrix Inversion Lemma [27]): If $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{C} \in \mathbb{R}^{n \times n}$ are non-singular, $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{D} \in \mathbb{R}^{m \times n}$,

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B} \left(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1}\right)^{-1} \mathbf{D}\mathbf{A}^{-1} \tag{28}$$

Proof: Now we state the proof of Theorem 3.1. Note that when the kernel bandwidth $\sigma \to \infty$,

$$\begin{aligned}
\hat{l}_{\mathbf{R}_k} &= \mathcal{G}_\sigma \left(\|\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{m}}_k)\|_{\mathbf{R}_k}\right) \\
&= \exp\left(\frac{-(\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{m}}_k))^\top \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{m}}_k))}{2\sigma^2}\right) \\
&\to 1,
\end{aligned} \tag{29}$$

Similarly, one can conclude that $\hat{l}_{\hat{\mathbf{C}}_k} \to 1$ when $\sigma \to \infty$. These arguments imply $\tilde{\mathbf{K}}_k$ in (22) with the approximation (23) become

$$\tilde{\mathbf{K}}_k = \left(\hat{\mathbf{C}}_k^{-1} + \mathbf{H}_k^\top \mathbf{R}_k^{-1}\mathbf{H}_k\right)^{-1} \mathbf{H}_k^\top \mathbf{R}_k^{-1}, \tag{30}$$

Then in view of Lemma 1.1 (we choose $\mathbf{A} = \hat{\mathbf{C}}_k^{-1}$, $\mathbf{B} = \mathbf{H}_k^\top$, $\mathbf{C} = \mathbf{R}_k$ and $\mathbf{D} = \mathbf{H}_k$), one can conclude that

$$\begin{aligned}
\tilde{\mathbf{K}}_k &= \left(\hat{\mathbf{C}}_k^{-1} + \mathbf{H}_k^\top \mathbf{R}_k^{-1}\mathbf{H}_k\right)^{-1} \mathbf{H}_k^\top \mathbf{R}_k^{-1} \\
&= \left(\hat{\mathbf{C}}_k - \hat{\mathbf{C}}_k\mathbf{H}_k^\top \left(\mathbf{R}_k + \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right)^{-1} \mathbf{H}_k\hat{\mathbf{C}}_k\right) \mathbf{H}_k^\top \mathbf{R}_k^{-1} \\
&= \hat{\mathbf{C}}_k\mathbf{H}_k^\top \mathbf{R}_k^{-1} - \hat{\mathbf{C}}_k\mathbf{H}_k^\top \left(\mathbf{R}_k + \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right)^{-1} \\
&\quad \times \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top \mathbf{R}_k^{-1} \\
&= \hat{\mathbf{C}}_k\mathbf{H}_k^\top \left(\mathbf{R}_k^{-1} - \left(\mathbf{R}_k + \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right)^{-1} \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top \mathbf{R}_k^{-1}\right) \\
&= \hat{\mathbf{C}}_k\mathbf{H}_k^\top \left(\mathbf{I} - \left(\mathbf{R}_k + \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right)^{-1} \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right) \mathbf{R}_k^{-1}.
\end{aligned} \tag{31}$$

Now use the Lemma 1.1 again (we choose $\mathbf{A} = \mathbf{I}$, $\mathbf{B} = \mathbf{I}$, $\mathbf{C} = \mathbf{R}_k$ and $\mathbf{D} = \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top$), one can further conclude that

$$\begin{aligned}
\tilde{\mathbf{K}}_k &= \hat{\mathbf{C}}_k\mathbf{H}_k^\top \left(\mathbf{I} - \left(\mathbf{R}_k + \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right)^{-1} \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right) \mathbf{R}_k^{-1} \\
&= \hat{\mathbf{C}}_k\mathbf{H}_k^\top \left(\mathbf{I} + \mathbf{R}_k^{-1}\mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right)^{-1} \mathbf{R}_k^{-1} \\
&= \hat{\mathbf{C}}_k\mathbf{H}_k^\top \left(\mathbf{R}_k + \mathbf{H}_k\hat{\mathbf{C}}_k\mathbf{H}_k^\top\right)^{-1},
\end{aligned} \tag{32}$$

which is equal to (8). ∎

REFERENCES

[1] S. Särkkä, *Bayesian filtering and smoothing.* Cambridge university press, 2013, no. 3.
[2] J. D. Hamilton, "State-space models," *Handbook of econometrics*, vol. 4, pp. 3039–3080, 1994.
[3] S. Chen, "Kalman filter for robot vision: a survey," *IEEE Transactions on industrial electronics*, vol. 59, no. 11, pp. 4409–4420, 2011.
[4] K. Law, A. Stuart, and K. Zygalakis, "Data assimilation," *Cham, Switzerland: Springer*, vol. 214, p. 52, 2015.
[5] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
[6] M. Hoshiya, E. Saito *et al.*, "Structural identification by extended kalman filter," *Jour. of Eng. Mech., ASCE*, vol. 110, no. 12, 1984.
[7] S. J. Julier and J. K. Uhlmann, "New extension of the kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, vol. 3068. International Society for Optics and Photonics, 1997, pp. 182–193.
[8] I. Arasaratnam and S. Haykin, "Cubature kalman filters," *IEEE Transactions on automatic control*, vol. 54, no. 6, pp. 1254–1269, 2009.
[9] G. Evensen, "The ensemble kalman filter: Theoretical formulation and practical implementation," *Ocean dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
[10] Y. Lu, B. Li, H. R. Karimi, and N. Zhang, "Measurement outlier-resistant target tracking in wireless sensor networks with energy harvesting constraints," *Journal of the Franklin Institute*, 2022.
[11] Y. Wang, Y. Sun, V. Dinavahi, S. Cao, and D. Hou, "Adaptive robust cubature kalman filter for power system dynamic state estimation against outliers," *IEEE Access*, vol. 7, pp. 105 872–105 881, 2019.
[12] A. H. Jazwinski, *Stochastic processes and filtering theory.* Courier Corporation, 2007.
[13] Z. Wu, J. Shi, X. Zhang, W. Ma, B. Chen, and I. Senior Member, "Kernel recursive maximum correntropy," *Signal Processing*, vol. 117, pp. 11–16, 2015.
[14] J. C. Principe, *Information theoretic learning: Renyi's entropy and kernel perspectives.* Springer Science & Business Media, 2010.
[15] V. Stojanovic and N. Nedic, "Robust kalman filtering for nonlinear multivariable stochastic systems in the presence of non-gaussian noise," *International Journal of Robust and Nonlinear Control*, vol. 26, no. 3, pp. 445–460, 2016.
[16] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust student's t-based kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545–1554, 2017.
[17] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on signal processing*, vol. 55, no. 11, pp. 5286–5298, 2007.
[18] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.
[19] G. Wang, N. Li, and Y. Zhang, "Maximum correntropy unscented kalman and information filters for non-gaussian measurement noise," *Journal of the Franklin Institute*, vol. 354, no. 18, pp. 8659–8677, 2017.
[20] G. Wang, Y. Zhang, and X. Wang, "Iterated maximum correntropy unscented kalman filters for non-gaussian systems," *Signal Processing*, vol. 163, pp. 87–94, 2019.
[21] Y. Tao and S. S.-T. Yau, "Outlier-robust iterative extended kalman filtering," *IEEE Signal Processing Letters*, vol. 30, pp. 743–747, 2023.
[22] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA journal*, vol. 3, no. 8, pp. 1445–1450, 1965.
[23] Y. Feng, X. Huang, L. Shi, Y. Yang, J. A. Suykens *et al.*, "Learning with the maximum correntropy criterion induced losses for regression." *J. Mach. Learn. Res.*, vol. 16, no. 30, pp. 993–1034, 2015.
[24] M. Zupanski, "Maximum likelihood ensemble filter: Theoretical aspects," *Monthly Weather Review*, vol. 133, no. 6, pp. 1710–1726, 2005.
[25] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
[26] B. Chen, L. Xing, H. Zhao, N. Zheng, J. C. Prı *et al.*, "Generalized correntropy for robust adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3376–3387, 2016.
[27] W. W. Hager, "Updating the inverse of a matrix," *SIAM review*, vol. 31, no. 2, pp. 221–239, 1989.