Multi-Kernel Maximum Correntropy Kalman Filter

Shilei Li, Dawei Shi, *Senior Member, IEEE*, Wulin Zou, and Ling Shi, *Senior Member, IEEE*

***Abstract*—Maximum correntropy criterion (MCC) has been widely used in Kalman filter to cope with heavy-tailed measurement noises. However, its performance on mitigating non-Gaussian process noises and unknown disturbance is rarely explored. In this letter, we extend the definition of correntropy from a single kernel to multiple kernels. Then, we derive a multi-kernel maximum correntropy Kalman filter (MKMCKF) to cope with multivariate non-Gaussian noises and disturbance. Three examples are provided to show the effectiveness of the proposed methods.**

***Index Terms*—Multi-kernel correntropy, Kalman filter, non-Gaussian noises, disturbance.**

I. Introduction

CORRENTROPY is a local similarity measure of two random variables in the kernel space where the kernel bandwidth acts as a zoom lens of controlling the "observation window" [1]. Naturally, it is not sensitive to outliers and is used as a robust cost for state estimation [2]–[6], adaptive filtering [7], regression and machine learning [8].

Kalman filter (KF) is a linear estimator that provides an optimal recursive solution for linear systems. It has been widely used in the field of navigation, target tracking, sensor fusion and many others. Nevertheless, its performance may degenerate with non-Gaussian noises and unknown disturbance. To handle this problem, Chen *et al.* [2] derive a maximum correntropy Kalman filter (MCKF) under MCC to cope with heavy-tailed measurement noises. After that, many variants or extensions of KF, such as Chandrasekhar-based Kalman filter [3], extended Kalman filter (EKF) [4], cubature Kalman filter (CKF) [5], unscented Kalman filter (UKF) [6], distributed Kalman filter (DKF) [9], and interacting multiple model Kalman filter (IMMKF) [10], have been re-derived under MCC. These filters, in general, are robust to measurement outliers. However, they may perform poorly with non-Gaussian process noises and unknown disturbance. One reason is that these filters utilize a unified kernel bandwidth to handle multivariate non-Gaussian noises. Another reason is that disturbance is difficult to be modeled accurately.

In some applications, systems are contaminated by non-Gaussian process noises or unknown disturbance, such as impulsive forces in manipulator collision [11], unknown input in maneuvering target tracking [12], and measurement disturbance in accelerometer for orientation estimation [13]. A conventional way to handle this problem is to increase the process covariance or measurement covariance in KF. However, this will corrupt the estimate accuracy with the disappearance of non-Gaussian noises or disturbance. A popular way to suppress disturbance is to augment the state and construct an augmented state Kalman filter (ASKF). Nevertheless, its performance is usually unsatisfactory since disturbance dynamics is unknown. To handle the aforementioned issues, Aravkin *et al.* [14] utilize the property of ℓ_1 norm to mitigate outliers or process disturbance in terms of smoothing, while Goel and Hassibi [15] design a regret-optimal estimator to suppress disturbance with robust estimation technology. Liu *et al.* [1] report that correntropy is closely related with a correntropy induced metric (CIM) between two random variables X and Y satisfying all properties of a metric. The scale of the norm in CIM can be adjusted among ℓ_0, ℓ_1 and ℓ_2 norm according to the Gaussian kernel bandwidth. In this letter, we extend the original correntropy from a single kernel to multiple kernels. Then we provide a weighted maximum correntropy criterion where the weights act as a tradeoff between the process fit and the measurement fit. Finally, we derive a novel MKMCKF that can cope with multivariate non-Gaussian noises and disturbance effectively. We tune the kernel bandwidths by employing the properties of CIM so that the non-Gaussian channels are analogical to an ℓ_1 or ℓ_0 norm while the Gaussian channels are analogical to an ℓ_2 norm in our objective function. The main contributions of this letter are summarized as follows.

1) We extend the definition of correntropy from a single kernel to multiple kernels. Then, we derive a novel MKMCKF that is robust to non-Gaussian noises and disturbance.

Manuscript received July 16, 2021; revised August 26, 2021; accepted September 13, 2021. Date of publication September 21, 2021; date of current version September 28, 2021. This work was supported by Shenzhen–Hong Kong Technology Cooperation Funding Scheme under Grant GHP-001-18SZ. Recommended by Senior Editor J.-F. Zhang. *(Corresponding author: Shilei Li.)*

Shilei Li and Ling Shi are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: slidk@connect.ust.hk; eesling@ust.hk).

Dawei Shi is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: daweishi@bit.edu.cn).

Wulin Zou is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, and also with the Control Department, Xeno Dynamics Company Ltd., Shenzhen 518000, China (e-mail: wzouab@connect.ust.hk).

Digital Object Identifier 10.1109/LCSYS.2021.3114137

2) We reveal that the performance of MKMCKF will degenerate with a special dynamic form. Then, we recover it by reconstructing system dynamics.
3) We show that MKMCKF is remarkably better than ASKF, MCKF and particle filter (PF) with non-Gaussian noises and disturbance in simulation.

The reminder is organized as follows. In Section II, we introduce the multi-kernel correntropy and system dynamics. In Section III, we derive the MKMCKF under MCC. In Section IV, we utilize MKMCKF for disturbance mitigation. In Section V, we give three examples to demonstrate the effectiveness of the proposed methods. In Section VI, we draw a conclusion.

II. Preliminaries

A. Correntropy

Correntropy is originally defined to measure the local similarity of two random variables. Given two random variables $X, Y \in \mathcal{R}$ with joint distribution $F_{XY}(x, y)$, the correntropy is given as

$$C(X, Y) = \mathrm{E}[\kappa(X, Y)] = \int \kappa(x, y) dF_{XY}(x, y) \tag{1}$$

where $\kappa(x, y)$ is a shift-invariant Mercer kernel. It is usually set to be Gaussian Kernel with

$$\kappa(x, y) = \mathrm{G}_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right) \tag{2}$$

where $e = x - y$ is the error and $\sigma > 0$ is the kernel bandwidth.

In this work, we extend the definition of correntropy from random variables to random vectors. Given two random vectors $\mathcal{X}, \mathcal{Y} \in \mathcal{R}^p$, the correntropy is defined as a weighted sum correntropy of each element:

$$\mathcal{C}(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^{p} \mathrm{E}[\bar{\kappa}(\mathcal{X}_i, \mathcal{Y}_i)] = \sum_{i=1}^{p} \int \bar{\kappa}(\mathrm{x}_i, \mathrm{y}_i) dF_{\mathcal{X}_i \mathcal{Y}_i}(\mathrm{x}_i, \mathrm{y}_i) \tag{3}$$

with

$$\bar{\kappa}(\mathrm{x}_i, \mathrm{y}_i) = \sigma_i^2 \exp\left(-\frac{\mathrm{e}_i^2}{2\sigma_i^2}\right) \tag{4}$$

where $\bar{\kappa}(\mathrm{x}_i, \mathrm{y}_i)$ is the Gaussian function, σ_i is the kernel bandwidth for i-th element, and $\mathrm{e}_i = \mathrm{x}_i - \mathrm{y}_i$ is the i-th error. It can be seen that (3) defines a new local similarity metric for two random vectors where the kernel vector $\sigma = [\sigma_1, \sigma_2, \ldots, \sigma_p]'$ controls both the "observation window" and the weighting coefficient.

In many applications, the samples are limited and the joint distribution is unknown. Then, equation (3) can be simplified as

$$\mathcal{C}(X, Y) = \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{p} \sigma_i^2 \exp\left(-\frac{\mathrm{e}_i^2(k)}{2\sigma_i^2}\right) \tag{5}$$

where $\mathrm{e}_i(k) = \mathrm{x}_i(k) - \mathrm{y}_i(k)$ is the error for element i at the k-th sampling, and N is the sample numbers.

B. System Model

We consider the following linear model:

$$\begin{aligned} \mathrm{x}_{k+1} &= \Phi \mathrm{x}_k + \mathrm{q}_k \\ \mathrm{y}_k &= \mathrm{H}\mathrm{x}_k + \mathrm{r}_k \end{aligned} \tag{6}$$

where $\mathrm{x}_k \in \mathcal{R}^n$ is the state, $\mathrm{y}_k \in \mathcal{R}^m$ is the measurement, Φ is the state transition matrix, H is the observation matrix, and q_k and r_k are white noises which are assumed to be independent with each other. Their covariance matrices are

$$\mathrm{Q}_k = \mathrm{E}[\mathrm{q}_k \mathrm{q}_k'], \quad \mathrm{R}_k = \mathrm{E}[\mathrm{r}_k \mathrm{r}_k']. \tag{7}$$

Traditional MCKF provides an tool for suppressing non-Gaussian measurement noises. However, its performance may degenerate with multivariate non-Gaussian process noises. To address this issue, we derive a novel filter under our multi-kernel correntropy.

C. Derivation of Algorithm

For a linear system described in (6), it can be rewritten as

$$\begin{pmatrix} \mathrm{x}_k^- \\ \mathrm{y}_k \end{pmatrix} = \begin{pmatrix} \mathrm{I} \\ \mathrm{H} \end{pmatrix} \mathrm{x}_k + \nu_k \tag{8}$$

where $\mathrm{I} \in \mathbb{R}^{n \times n}$ is an identity matrix and x_k^- is the *a priori* estimate of state x_k. The noise term ν_k is

$$\nu_k = \begin{pmatrix} \mathrm{x}_k^- - \mathrm{x}_k \\ \mathrm{r}_k \end{pmatrix} \tag{9}$$

with

$$E(\nu_k \nu_k') = \begin{pmatrix} \mathrm{P}_k^- & 0 \\ 0 & \mathrm{R}_k \end{pmatrix} = \begin{pmatrix} \mathrm{B}_p \mathrm{B}_p' & 0 \\ 0 & \mathrm{B}_r \mathrm{B}_r' \end{pmatrix} = \mathrm{B}_k \mathrm{B}_k' \tag{10}$$

where P_k^- is the *a priori* error covariance while B_p and B_r can be obtained by Cholesky decomposition. Left multiplying B_k^{-1} in both sides of (8), we obtain

$$\mathrm{T}_k = \mathrm{W}_k \mathrm{x}_k + \xi_k \tag{11}$$

where

$$\mathrm{T}_k = \mathrm{B}_k^{-1} \begin{pmatrix} \mathrm{x}_k^- \\ \mathrm{y}_k \end{pmatrix}, \ \mathrm{W}_k = \mathrm{B}_k^{-1} \begin{pmatrix} \mathrm{I} \\ \mathrm{H} \end{pmatrix}, \ \xi_k = B_k^{-1} \nu_k \tag{12}$$

The noise term ξ_k is white since $\mathrm{E}(\xi_k \xi_k^T) = \mathrm{I}$.

In this letter, we present the following novel MCC-based objective function:

$$\begin{aligned} J(\mathrm{x}_k) &= \sum_{i=1}^{n+m} \sigma_i^2 \mathrm{G}_{\sigma_i}(\mathrm{e}_{i,k}) \\ &= \sum_{i=1}^{n+m} \sigma_i^2 \exp\left(-\frac{(\mathrm{t}_{i,k} - \mathrm{w}_{i,k}\mathrm{x}_k)^2}{2\sigma_i^2}\right) \end{aligned} \tag{13}$$

where $\mathrm{G}_{\sigma_i}(\mathrm{e}_{i,k})$ represents Gaussian function value of $\mathrm{e}_{i,k}$ with kernel bandwidth σ_i, $\mathrm{e}_{i,k} = \mathrm{t}_{i,k} - \mathrm{w}_{i,k}\mathrm{x}_k$ is the error at time step k, $\mathrm{t}_{i,k}$ is the i-th element of T_k, and $\mathrm{w}_{i,k}$ is the i-th row of W_k. Then, under MCC, the optimal estimator is

$$\mathrm{x}_k^+ = \arg\max_{\mathrm{x}_k} J(\mathrm{x}_k) = \arg\max_{\mathrm{x}_k} \sum_{i=1}^{n+m} \sigma_i^2 G_{\sigma_i}(\mathrm{e}_{i,k}). \tag{14}$$

The optimal solution can be obtained by solving

$$\frac{\partial J(\mathrm{x}_k)}{\partial \mathrm{x}_k} = 0. \tag{15}$$

It follows that

$$\sum_{i=1}^{n+m} G_{\sigma_i}(\mathrm{e}_{i,k})\mathrm{w}'_{i,k}\left(\mathrm{t}_{i,k} - \mathrm{w}_{i,k}\mathrm{x}_k\right) = 0. \tag{16}$$

Then, we have

$$\mathrm{x}_k = \left(\sum_{i=1}^{n+m} G_{\sigma_i}(\mathrm{e}_{i,k})\mathrm{w}'_{i,k}\mathrm{w}_{i,k}\right)^{-1}\left(\sum_{i=1}^{n+m} G_{\sigma_i}(\mathrm{e}_{i,k})\mathrm{w}'_{i,k}\mathrm{t}_{i,k}\right). \tag{17}$$

Since $\mathrm{e}_{i,k} = \mathrm{t}_{i,k} - \mathrm{w}_{i,k}\mathrm{x}_k$, both sides of (17) contains x_k which actually is a fixed-point function. It can be rewritten as

$$\mathrm{x}_k = (\mathrm{W}'_k\mathrm{M}_k\mathrm{W}_k)^{-1}\mathrm{W}'_k\mathrm{M}_k\mathrm{T}_k \tag{18}$$

with

$$\begin{aligned} \mathrm{M}_k &= \begin{pmatrix} \mathrm{M}_p & 0 \\ 0 & \mathrm{M}_r \end{pmatrix} \\ \mathrm{M}_p &= \mathrm{diag}[G_{\sigma_p}(\mathrm{e}_p)],\ \ \mathrm{M}_r = \mathrm{diag}[G_{\sigma_r}(\mathrm{e}_r)] \\ \mathrm{e}_p &= \mathrm{B}_p^{-1}(\mathrm{x}_k^- - \mathrm{x}_k),\ \ \mathrm{e}_r = \mathrm{B}_r^{-1}(\mathrm{y}_k - \mathrm{H}\mathrm{x}_k) \end{aligned} \tag{19}$$

where $\sigma_p = [\sigma_1, \sigma_2, \ldots, \sigma_n]'$ is the process bandwidth vector, $\sigma_r = [\sigma_{n+1}, \sigma_{n+2}, \ldots, \sigma_{n+m}]'$ is the measurement bandwidth vector, e_p is the process error, and e_r is the measurement error. Substituting (10) and (12) into (18), we arrive

$$(\mathrm{W}'_k\mathrm{M}_k\mathrm{W}_k)^{-1} = [(\mathrm{B}_p^{-1})'\mathrm{M}_p\mathrm{B}_p^{-1} + \mathrm{H}'(\mathrm{B}_r^{-1})'\mathrm{M}_r\mathrm{B}_r^{-1}\mathrm{H}]^{-1}. \tag{20}$$

Then, using matrix inversion lemma, we obtain

$$\begin{aligned} (\mathrm{W}'_k\mathrm{M}_k\mathrm{W}_k)^{-1} &= \mathrm{B}_p\mathrm{M}_p^{-1}\mathrm{B}'_p - \mathrm{B}_p\mathrm{M}_p^{-1}\mathrm{B}'_p\mathrm{H}' \\ &\times (\mathrm{B}_r\mathrm{M}_r^{-1}\mathrm{B}'_r + \mathrm{H}\mathrm{B}_p\mathrm{M}_p^{-1}\mathrm{B}'_p\mathrm{H}')^{-1}\mathrm{H}\mathrm{B}_p\mathrm{M}_p^{-1}\mathrm{B}'_p. \end{aligned} \tag{21}$$

Further, we have

$$\mathrm{W}'_k\mathrm{M}_k\mathrm{T}_k = (\mathrm{B}_p^{-1})'\mathrm{M}_p\mathrm{B}_p^{-1}\mathrm{x}_k^- + \mathrm{H}'(\mathrm{B}_r^{-1})'\mathrm{M}_r\mathrm{B}_r^{-1}\mathrm{y}_k. \tag{22}$$

Substituting (21) and (22) into (18), we obtain

$$\mathrm{x}_k = \mathrm{x}_k^- + \tilde{\mathrm{K}}(\mathrm{y}_k - \mathrm{H}\mathrm{x}_k^-) \tag{23}$$

with

$$\begin{aligned} \tilde{\mathrm{K}} &= \tilde{\mathrm{P}}_k^-\mathrm{H}'(\mathrm{H}\tilde{\mathrm{P}}_k^-\mathrm{H}' + \tilde{\mathrm{R}}_k)^{-1} \\ \tilde{\mathrm{P}}_k^- &= \mathrm{B}_p\mathrm{M}_p^{-1}\mathrm{B}'_p,\ \ \tilde{\mathrm{R}}_k = \mathrm{B}_r\mathrm{M}_r^{-1}\mathrm{B}'_r. \end{aligned} \tag{24}$$

Finally, the *a posteriori* error covariance can be updated with

$$\mathrm{P}_k^+ = (\mathrm{I} - \tilde{\mathrm{K}}_k\mathrm{H})\mathrm{P}_k^-(\mathrm{I} - \tilde{\mathrm{K}}_k\mathrm{H})' + \tilde{\mathrm{K}}_k\mathrm{R}_k\tilde{\mathrm{K}}'_k. \tag{25}$$

Notably, (23) is a fixed-point equation since $\tilde{\mathrm{P}}_k^-$ and $\tilde{\mathrm{R}}_k$ are related with x_k. Hence, a fixed-point algorithm can be utilized to solve it. To avoid the singularity of M_p and M_r, we introduce the following constraints:

$$\mathrm{M}_p \succcurlyeq \alpha\mathrm{I}_{n\times n},\ \ \mathrm{M}_r \succcurlyeq \beta\mathrm{I}_{m\times m} \tag{26}$$

where α and β are two small positive numbers. In a practical application, if the element of diagonal matrix M_p and M_r is smaller than α or β, it will be replaced by the scalar α or β correspondingly. Then, the fixed-point equation (23) will stop iteration. The detailed algorithm is summarized in Algorithm 1.

Algorithm 1 MKMCKF

1: **Step 1: Initialization**
2: Choose bandwidth vectors σ_p, σ_r, lower bound values α, β, and a threshold ε.
3: **Step 2: State Prediction**
4: $\hat{\mathrm{x}}_k^- = \Phi\hat{\mathrm{x}}_{k-1}^+$
5: $\mathrm{P}_k^- = \Phi\mathrm{P}_{k-1}^+\Phi' + \mathrm{Q}_k$
6: Obtain B_p with $\mathrm{P}_k^- = \mathrm{B}_p\mathrm{B}'_p$
7: Obtain B_r with $\mathrm{R}_k = \mathrm{B}_r\mathrm{B}'_r$
8: **Step 3: State Update**
9: $\hat{\mathrm{x}}_{k,0}^+ = \hat{\mathrm{x}}_k^-$
10: **while** $\frac{\left\|\hat{\mathrm{x}}_{k,t}^+ - \hat{\mathrm{x}}_{k,t-1}^+\right\|}{\left\|\hat{\mathrm{x}}_{k,t}^+\right\|} > \varepsilon$, $\mathrm{M}_p \succcurlyeq \alpha\mathrm{I}_{n\times n}$, $\mathrm{M}_r \succcurlyeq \beta\mathrm{I}_{m\times m}$ **do**
11: $\hat{\mathrm{x}}_{k,t}^+ = \hat{\mathrm{x}}_k^- + \tilde{\mathrm{K}}_{k,t}(\mathrm{y}_k - \mathrm{H}\hat{\mathrm{x}}_k^-)$ ▷ t starts from 1
12: $\tilde{\mathrm{K}}_{k,t} = \tilde{\mathrm{P}}_k^-\mathrm{H}'(\mathrm{H}\tilde{\mathrm{P}}_k^-\mathrm{H}' + \tilde{\mathrm{R}}_k)^{-1}$
13: $\tilde{\mathrm{P}}_k^- = \mathrm{B}_p\tilde{\mathrm{M}}_p^{-1}\mathrm{B}'_p$
14: $\tilde{\mathrm{R}}_k = \mathrm{B}_r\tilde{\mathrm{M}}_r^{-1}\mathrm{B}'_r$
15: $\tilde{\mathrm{M}}_p = \mathrm{diag}(G_{\sigma_p}(\mathrm{e}_p))$
16: $\tilde{\mathrm{M}}_r = \mathrm{diag}(G_{\sigma_r}(\mathrm{e}_r))$
17: $\mathrm{e}_p = \mathrm{B}_p^{-1}\hat{\mathrm{x}}_k^- - \mathrm{B}_p^{-1}\hat{\mathrm{x}}_{k,t-1}^+$
18: $\mathrm{e}_r = \mathrm{B}_r^{-1}\mathrm{y}_k - \mathrm{B}_r^{-1}\mathrm{H}\hat{\mathrm{x}}_{k,t-1}^+$
19: $t = t + 1$
20: **end while**
21: $\mathrm{P}_k^+ = (\mathrm{I} - \tilde{\mathrm{K}}_k\mathrm{H})\mathrm{P}_k^-(\mathrm{I} - \tilde{\mathrm{K}}_k\mathrm{H})' + \tilde{\mathrm{K}}_k\mathrm{R}_k\tilde{\mathrm{K}}'_k$

Remark 1: We use a threshold ε and constraints (26) as stop condition. This can avoid matrix singularity in numerical calculation.

Remark 2: MKMCKF can suppress both Gaussian noises and non-Gaussian noises. Hence, the selection of bandwidth vectors should balance between Gaussian noises elimination and non-Gaussian noises mitigation.

Theorem 1: MKMCKF is identical to tradition KF if we set all bandwidths $\sigma_i \to \infty$ and is equal to MCKF if we set all bandwidths to be the same with $\sigma_i = \sigma$.

Proof: As all $\sigma_i \to \infty$, we have $\mathrm{M}_k = \mathrm{I}$. Then, MKMCKF is equal to KF. As $\sigma_i = \sigma$, we have $\mathrm{M}_p = \mathrm{diag}[G_\sigma(\mathrm{e}_p)]$, $\mathrm{M}_r = \mathrm{diag}[G_\sigma(\mathrm{e}_r)]$. Then, it is identical to MCKF. ■

Theorem 2: Suppose that the measurement matrix contains l column zero vectors with $\mathrm{H} = [\mathrm{C}_{m\times f}, 0_{m\times l}]$ where $n = f + l$. Then, we write the process error e_p as $\mathrm{e}_p = [\mathrm{e}'_f, \mathrm{e}'_l]'$. The part e_l is always equal to a zero vector in MKMCKF.

Proof: In MKMCKF, the effects of non-Gaussian process noises are mitigated by the adaptive term $\tilde{\mathrm{P}}_k^-$ which is only related with e_p. Based on line 17 in Algorithm 1, we have

$$\mathrm{e}_p = \mathrm{B}_p^{-1}\hat{\mathrm{x}}_k^- - \mathrm{B}_p^{-1}\hat{\mathrm{x}}_{k,t-1}^+ \tag{27}$$

where t is the iteration number at time step k. If $t = 1$, we know that $\hat{\mathrm{x}}_{k,0}^+ = \hat{\mathrm{x}}_k^-$. Then, e_p is a zero vector and $\tilde{\mathrm{P}}_k^- = \mathrm{P}_k^-$. If $t > 1$, based on line 6, 11, 12, 17 in Algorithm 1, we have

$$\begin{aligned} \mathrm{e}_p &= \mathrm{B}_p^{-1}(\hat{\mathrm{x}}_k^- - \hat{\mathrm{x}}_{k,t-1}^+) \\ &= -\mathrm{B}_p^{-1}\tilde{\mathrm{K}}_{k,t-1}(\mathrm{y}_k - \mathrm{H}\hat{\mathrm{x}}_k^-) \\ &= -\mathrm{B}_p^{-1}\tilde{\mathrm{P}}_k^-\mathrm{H}'(\mathrm{H}\tilde{\mathrm{P}}_k^-\mathrm{H}' + \tilde{\mathrm{R}}_k)^{-1}(\mathrm{y}_k - \mathrm{H}\hat{\mathrm{x}}_k^-) \\ &= -\tilde{\mathrm{M}}_p^{-1}\mathrm{B}'_p\mathrm{H}'(\mathrm{H}\tilde{\mathrm{P}}_k^-\mathrm{H}' + \tilde{\mathrm{R}}_k)^{-1}(\mathrm{y}_k - \mathrm{H}\hat{\mathrm{x}}_k^-) \end{aligned} \tag{28}$$

Since $\tilde{\mathrm{M}}_p^{-1}$ is diagonal, B'_p is an upper triangular matrix, and H' contains l row zero vectors, it follows that $\tilde{\mathrm{M}}_p^{-1}\mathrm{B}'_p\mathrm{H}'$ can

be written as a block matrix with

$$\tilde{\mathrm{M}}_p^{-1}\mathrm{B}_p'\mathrm{H}' = \begin{pmatrix} \Lambda_{f\times m} \\ 0_{l\times m} \end{pmatrix}. \tag{29}$$

We know that $\eta = (\mathrm{H}\tilde{\mathrm{P}}_k^{-}\mathrm{H}' + \tilde{\mathrm{R}}_k)^{-1}(\mathrm{y}_k - \mathrm{H}\hat{\mathrm{x}}_k^{-}) \in \mathcal{R}^{m\times 1}$ is a vector. Then, we arrive

$$\mathrm{e}_p = \begin{pmatrix} \mathrm{e}_f \\ \mathrm{e}_l \end{pmatrix} = \begin{pmatrix} \Lambda_{f\times m} \\ 0_{l\times m} \end{pmatrix}\eta = \begin{pmatrix} \mathrm{e}_f \\ 0_l \end{pmatrix} \tag{30}$$

where e_p always contains l zero elements which correspond to measurement matrix part $0_{m\times l}$. The statement is proved. ■

Remark 3: In the situation of Theorem 2, we can write process kernel vector as $\sigma_p = [\sigma_f', \sigma_l']'$. Then, the selection of σ_l becomes meaningless since we always have $\mathrm{e}_l = 0$. With this special dynamic form, the performance of MKMCKF will degenerate due to the "singularity" of process error. To avoid this situation, we can rewrite system dynamics or using other decomposition method (such as principal square root decomposition) in Algorithm 1.

III. Disturbance Mitigation

In many applications, systems contain unknown process disturbance or measurement disturbance. These disturbances are usually difficult to be estimated due to the lack of accurate disturbance models, hence contaminating the estimate accuracy. In this section, we employ MKMCKF to handle this problem.

A. Systems With Unknown Process Disturbance

We consider the following linear model with unknown process disturbance:

$$\begin{aligned} \mathrm{x}_{k+1} &= \Phi\mathrm{x}_k + \Gamma\mathrm{d}_k + \mathrm{q}_k \\ \mathrm{y}_k &= \mathrm{H}\mathrm{x}_k + \mathrm{r}_k \end{aligned} \tag{31}$$

where d_k is the unknown disturbance, matrix Γ maps the disturbance to the state, and q_k and r_k are Gaussian noises.

To eliminate the influence of d_k on state x_k, we construct an ASKF and augment the state as $\bar{\mathrm{x}}_k = [\mathrm{x}_k', \mathrm{d}_k']'$. Since we have no prior knowledge about disturbance, a conventional way is to assume that $\mathrm{d}_{k+1} = \mathrm{d}_k$. Correspondingly, the system dynamics can be written as

$$\begin{aligned} \bar{\mathrm{x}}_{k+1} &= \bar{\Phi}\bar{\mathrm{x}}_k + \bar{\mathrm{q}}_k \\ \mathrm{y}_k &= \bar{\mathrm{H}}\bar{\mathrm{x}}_k + \mathrm{r}_k \end{aligned} \tag{32}$$

with

$$\bar{\Phi} = \begin{bmatrix} \Phi & \Gamma \\ 0 & \mathrm{I} \end{bmatrix},\ \bar{\mathrm{H}} = \begin{bmatrix} \mathrm{H} & 0 \end{bmatrix} \tag{33}$$

where $\bar{\mathrm{q}}_k$ and r_k are Gaussian noises with covariance matrices

$$\mathrm{Q}_k = \mathrm{E}[\bar{\mathrm{q}}_k\bar{\mathrm{q}}_k'], \quad \mathrm{R}_k = \mathrm{E}[\mathrm{r}_k\mathrm{r}_k']. \tag{34}$$

Notably, the modeling of d_k is not accurate and the unmodeled dynamics can be seen as non-Gaussian process noises. According to Theorem 2, the performance of MKMCKF will degenerate due to the special form of measurement matrix $\bar{\mathrm{H}}$. To cope with this issue, we exchange the position of state and disturbance and write the new state as $\bar{\mathrm{x}}_k^* = [\mathrm{d}_k', \mathrm{x}_k']'$. Then, the system dynamics can be rewritten as

$$\begin{aligned} \bar{\mathrm{x}}_{k+1}^* &= \bar{\Phi}^*\bar{\mathrm{x}}_k^* + \mathrm{q}_k^* \\ \mathrm{y}_k &= \bar{\mathrm{H}}^*\bar{\mathrm{x}}_k^* + \mathrm{r}_k \end{aligned} \tag{35}$$

with

$$\bar{\Phi}^* = \begin{bmatrix} \mathrm{I} & 0 \\ \Gamma & \Phi \end{bmatrix},\ \bar{\mathrm{H}}^* = \begin{bmatrix} 0 & \mathrm{H} \end{bmatrix} \tag{36}$$

where q_k^* is the process noises. With this new equation, the situation of Theorem 2 can be avoided. A performance comparison of MKMCKF with equation (32) and (35) is shown in IV-A and in IV-B.

B. Systems With Unknown Measurement Disturbance

We consider the following linear system with unknown measurement disturbance:

$$\begin{aligned} \mathrm{x}_{k+1} &= \Phi x_k + \mathrm{q}_k \\ \mathrm{y}_k &= \mathrm{H}x_k + \Omega\mathrm{d}_k + \mathrm{r}_k \end{aligned} \tag{37}$$

where d_k is the unknown measurement disturbance and Ω maps the disturbance to the measurement. We construct an ASKF and write the new state as $\bar{\mathrm{x}}_k = [\mathrm{x}_k', \mathrm{d}_k']'$. We assume that the disturbance has a decayed first-order Markov model with a decay coefficient $0 < c < 1$. Then, we have

$$\begin{aligned} \bar{\mathrm{x}}_{k+1} &= \bar{\Phi}\bar{\mathrm{x}}_k + \bar{\mathrm{q}}_k \\ \mathrm{y}_k &= \bar{\mathrm{H}}\bar{\mathrm{x}}_k + \mathrm{r}_k \end{aligned} \tag{38}$$

with

$$\bar{\Phi} = \begin{bmatrix} \Phi & 0 \\ 0 & c\cdot\mathrm{I} \end{bmatrix},\ \bar{\mathrm{H}} = \begin{bmatrix} \mathrm{H} & \Omega \end{bmatrix}$$

where $\bar{\mathrm{q}}_k$ is the process noises and r_k is the measurement noises. Since the modeling of disturbance $\mathrm{d}_{k+1} = c\cdot\mathrm{d}_k$ is not accurate, the performance of ASKF will degenerate with the existence of measurement disturbance. However, this can be mitigated in MKMCKF by selecting kernels properly. Comparisons of ASKF, MCKF and MKMCKF and particle filter (PF) are shown in IV-C.

IV. Simulations

In this section, we give three examples to demonstrate the abilities of MKMCKF on mitigating non-Gaussian process noises, eliminating process disturbance, and suppressing measurement disturbance.

A. Example 1

We consider the following velocity tracking problem with non-Gaussian process noises:

$$\begin{aligned} \begin{bmatrix} \mathrm{x}_{1,k+1} \\ \mathrm{x}_{2,k+1} \end{bmatrix} &= \begin{bmatrix} 1 & \mathrm{T} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \mathrm{x}_{1,k} \\ \mathrm{x}_{2,k} \end{bmatrix} + \begin{bmatrix} \mathrm{q}_{1,k} \\ \mathrm{q}_{2,k} \end{bmatrix} \\ \mathrm{y}_k &= \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} \mathrm{x}_{1,k} \\ \mathrm{x}_{2,k} \end{bmatrix} + \mathrm{r}_k \end{aligned} \tag{39}$$

where k is time index, T is sampling time with $\mathrm{T} = 0.1\mathrm{s}$, and $\mathrm{x}_k = [\mathrm{x}_{1,k}, \mathrm{x}_{2,k}]'$ represents target velocity and acceleration at time step k. The process noises and measurement noises are given as

$$\begin{aligned} \mathrm{q}_{1,k} &\sim 0.9\mathcal{N}(0, 0.01) + 0.1\mathcal{N}(0, 4) \\ \mathrm{q}_{2,k} &\sim 0.9\mathcal{N}(0, 0.01) + 0.1\mathcal{N}(0, 100) \\ \mathrm{r}_k &\sim \mathcal{N}(0, 0.04). \end{aligned} \tag{40}$$

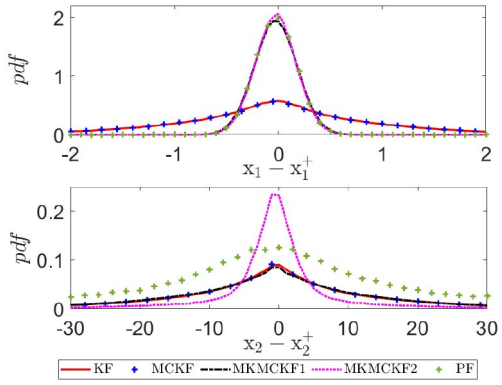

Fig. 1. Probability densities of estimation error with KF, MCKF, MKMCKF1, MKMCKF2 and PF. The kernel bandwidth for MCKF is $\sigma = 40$. The bandwidths for MKMCKF1 are set to be $\sigma_p = [1.2, 0.5]'$ and $\sigma_r = 10^4$, and the bandwidths for MKMCKF2 are $\sigma_p = [0.5, 1.2]'$ and $\sigma_r = 10^4$.

TABLE I
PERFORMANCES OF KF, MCKF, MKMCKF1, MKMCKF2 AND PF WITH NON-GAUSSIAN PROCESS NOISES

Filters	RMSE of $x_1(m/s)$	RMSE of $x_2(m/s^2)$
KF	1.072	8.377
MCKF	1.056	8.426
MKMCKF1	0.2050	8.355
MKMCKF2	0.1968	5.003
PF	0.2026	6.3108

Based on Theorem 2, MKMCKF with equation (39) cannot suppress non-Gaussian process noises effectively. To handle this problem, we exchange the position of $x_{1,k}$ and $x_{2,k}$ and write a new state as $x_k^* = [x_{2,k}, x_{1,k}]'$. Then, the system dynamics can be rewritten as

$$\begin{bmatrix} x_{2,k+1} \\ x_{1,k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ T & 1 \end{bmatrix} \begin{bmatrix} x_{2,k} \\ x_{1,k} \end{bmatrix} + \begin{bmatrix} q_{2,k} \\ q_{1,k} \end{bmatrix}$$
$$y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_{2,k} \\ x_{1,k} \end{bmatrix} + r_k. \tag{41}$$

Five tests are conducted to verify the performance of KF, MCKF, MKMCKF1, MKMCKF2 and PF. The former three filters use the dynamics equation (39) while the fourth employs (41). The initial state, process covariance, measurement covariance and initial error covariance are set to be the same for the former four filters. The particle number for PF is $N = 1000$ while the resampling method is systematic resample. As for the kernel bandwidths, we select them based on the property of CIM and tune them by trail and error in MKMCKF. A total of 500 independent Monte Carlo runs are conducted with 1000 time steps at each run. The error probability density functions of KF, MCKF, MKMCKF1, and MKMCKF2 are drawn in Fig. 1. The corresponding RMSE performance is summarized in Table I. It can be seen that MKMCKF2 outperforms other filters with non-Gaussian process noises, even including the heavy computation required PF. In addition, one can see that the performance of MKMCKF2 is better than MKMCKF1, which verifies Remark 3.

TABLE II
PERFORMANCE OF ASKF, MCKF, MKMCKF1, MKMCKF2 AND PF WITH PROCESS DISTURBANCE

Filter	RMSE of $x_1(m/s)$	RMSE of $x_2(N)$
ASKF	0.5077	3.6949
MCKF	0.5065	3.6916
MKMCKF1	0.1534	4.3119
MKMCKF2	0.1450	1.0390
PF	0.1715	1.2744

B. Example 2

We consider a velocity tracking problem while the target is occasionally manipulated by unknown maneuvering force f. The target mass is assumed to be $m = 1$ while the acceleration is calculated by $\frac{f}{m}$. We assume that $f_{k+1} = f_k$ and construct an ASKF by setting $x_k = [x_{1,k}, x_{2,k}]' = [v_k, f_k]'$ which represents the target velocity and the maneuvering force. Then, the system dynamics can be written as

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 1 & \frac{T}{m} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} q_{1,k} \\ q_{2,k} \end{bmatrix}$$
$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + r_k. \tag{42}$$

The process noises and measurement noises are both Gaussian with

$$q_{1,k} \sim \mathcal{N}(0, 0.01),\ q_{2,k} \sim \mathcal{N}(0, 0.01),\ r(k) \sim \mathcal{N}(0, 0.04). \tag{43}$$

We assume that the target is driven by a force with the following expression:

$$f_k(t) = \begin{cases} 20\sin(0.4\pi t) + r_k, & 50 \le t \le 60 \\ r_k, & otherwise \end{cases} \tag{44}$$

According to Theorem 2, the performance of MKMCKF will degenerate with $H = [1, 0]$. To handle this problem, we rewrite the system dynamics based on (35). Then, we have

$$\begin{bmatrix} x_{2,k+1} \\ x_{1,k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{T}{m} & 1 \end{bmatrix} \begin{bmatrix} x_{2,k} \\ x_{1,k} \end{bmatrix} + \begin{bmatrix} q_{2,k} \\ q_{1,k} \end{bmatrix}$$
$$y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_{2,k} \\ x_{1,k} \end{bmatrix} + r_k. \tag{45}$$

The new state is $x_k^* = [x_{2,k}, x_{1,k}]'$ and the situation of Theorem 2 is avoided. Five tests are conducted to verify the performance of ASKF, MCKF, MKMCKF1, MKMCKF2 and PF. The former three filters utilize equation (42) while the fourth uses (45). The process covariance, measurement covariance, initial state and initial error covariance are set to be the same for the former four filters. The particle number for PF is $N = 1000$ while the resampling method is systematic resample. The estimate error in a run is shown in Fig. 2. After 500 independent Monte Carlo runs, the RMSE performance is summarized in Table II. It can be seen that MKMCKF2 outperforms ASKF, MCKF, MKMCKF1 and PF remarkably on both the estimate of state $x_{1,k}$ and the disturbance $x_{2,k}$. One reason for the poor performance of PF may be that the state $x_{2,k}$ cannot be measured directly.

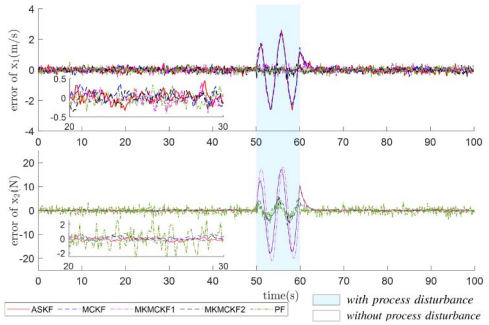

Fig. 2. Error performance of ASKF, MCKF, MKMCKF1, MKMCKF2 and PF. The white region is the error without process disturbance while the blue region is the error with process disturbance. The kernel bandwidth for MCKF is $\sigma = 100$ while the bandwidths for MKMCKF1 are $\sigma_p = [1.5, 10^4]'$ and $\sigma_r = 10^4$, and the bandwidths for MKMCKF2 are $\sigma_p = [10^4, 1.5]'$ and $\sigma_r = 10^4$. The zoom in plot at time interval $20s$ to $30s$ shows that the performance of ASKF, MCKF, MKMCKF1 and MKMCKF2 is similar without process disturbance.

TABLE III
PERFORMANCE OF ASKF, MCKF, MKMCKF AND PF WITH MEASUREMENT DISTURBANCE

Filter	RMSE of x_1(m/s)	RMSE of x_2(m/s)
ASKF	1.946	1.9925
MCKF	0.5809	2.2284
MKMCKF	0.4111	0.4120
PF	0.5739	0.7847

C. Example 3

We consider the following velocity tracking problem while the measurement is contaminated by unknown disturbance:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} q_{1,k} \\ q_{2,k} \end{bmatrix}$$
$$y_k = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + r_k \tag{46}$$

where $x_k = [x_{1,k}, x_{2,k}]' = [v_k, d_k]'$ represents target velocity and disturbance velocity, and c is a decay coefficient with $c = 0.8$. Notably, the measurement contains both the target velocity and the disturbance. The disturbance dynamics are assumed to be $d_{k+1} = c \cdot d_k$.

In simulation, the disturbance has the following expression:

$$d_k(t) = \begin{cases} 10\sin(0.4\pi t) + r_k, & 50 \le t \le 60 \\ r_k, & otherwise. \end{cases} \tag{47}$$

The initial state, initial error covariance, process covariance and measurement covariance are set to be the same for ASKF, MCKF and MKMCKF. Again, the particle number for PF is $N = 1000$ while the resampling method is systematic resample. The estimate error of these filters in a run is shown in Fig. 3. After 500 independent Monte Carlo runs, the RMSE error performance of these filters is summarized in Table III. It can be seen that though MCKF can estimate state x_1 accurately while its performance on disturbance x_2 is poor. On the contrary, MKMCKF can estimate both the state and the disturbance accurately.

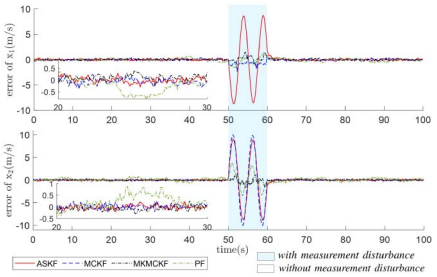

Fig. 3. Error performance of ASKF, MCKF, MKMCKF and PF with measurement disturbance. The white region represents the error without measurement disturbance while the blue region is the error with measurement disturbance. The kernel bandwidth for MCKF is $\sigma = 5$ while the bandwidths for MKMCKF are $\sigma_p = [10^4, 5]'$ and $\sigma_r = 10^4$. The zoom in plot shows that the performance of ASKF, MCKF, and MKMCKF is similar without disturbance.

V. CONCLUSION

In this letter, we extend the definition of correntropy to a multi-kernel version and derive a novel MKMCKF filter under MCC. It can handle multi-dimensional non-Gaussian noises effectively. Moreover, we demonstrate its abilities on disturbance mitigation. Simulations verify our methods.

REFERENCES

[1] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[2] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.
[3] M. V. Kulikova, "Chandrasekhar-based maximum correntropy Kalman filtering with the adaptive kernel size selection," *IEEE Trans. Autom. Control*, vol. 65, no. 2, pp. 741–748, Feb. 2020.
[4] X. Liu, Z. Ren, H. Lyu, Z. Jiang, P. Ren, and B. Chen, "Linear and nonlinear regression-based maximum correntropy extended Kalman filtering," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 5, pp. 3093–3102, May 2021.
[5] X. Liu, H. Qu, J. Zhao, and P. Yue, "Maximum correntropy square-root cubature Kalman filter with application to SINS/GPS integrated systems," *ISA Trans.*, vol. 80, pp. 195–202, Sep. 2018.
[6] X. Liu, B. Chen, B. Xu, Z. Wu, and P. Honeine, "Maximum correntropy unscented filter," *Int. J. Syst. Sci.*, vol. 48, no. 8, pp. 1607–1615, 2017.
[7] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *Proc. Int. Joint Conf. Neural Netw.*, Atlanta, GA, USA, 2009, pp. 2950–2955.
[8] B. Chen, X. Wang, N. Lu, S. Wang, J. Cao, and J. Qin, "Mixture correntropy for robust learning," *Pattern Recognit.*, vol. 79, pp. 318–327, Jul. 2018.
[9] G. Wang, R. Xue, and J. Wang, "A distributed maximum correntropy Kalman filter," *Signal Process.*, vol. 160, pp. 247–251, Jul. 2019.
[10] X. Fan, G. Wang, J. Han, and Y. Wang, "Interacting multiple model based on maximum correntropy Kalman filter," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 8, pp. 3017–3021, Aug. 2021.
[11] S. Kang, K. Komoriya, K. Yokoi, T. Koutoku, B. Kim, and S. Park, "Control of impulsive contact force between mobile manipulator and environment using effective mass and damping controls," *Int. J. Precis. Eng. Manuf.*, vol. 11, no. 5, pp. 697–704, 2010.
[12] Y. Bar-Shalom, K. Chang, and H. A. P. Blom, "Tracking a maneuvering target using input estimation versus the interacting multiple model algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 2, pp. 296–300, Mar. 1989.
[13] D. Roetenberg, H. J. Luinge, C. T. M. Baten, and P. H. Veltink, "Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 395–405, Sep. 2005.
[14] A. Aravkin, J. V. Burke, L. Ljung, A. Lozano, and G. Pillonetto, "Generalized Kalman smoothing: Modeling and algorithms," *Automatica*, vol. 86, pp. 63–86, Dec. 2017.
[15] G. Goel and B. Hassibi, "Regret-optimal estimation and control," 2021. [Online]. Available: arXiv:2106.12097.