Gaussian-Cauchy Mixture Kernel Function Based Maximum Correntropy Criterion Kalman Filter for Linear Non-Gaussian Systems

Quanbo Ge, *Member, IEEE*, Xuefei Bai, *Student Member, IEEE*, and Pingliang Zeng, *Senior Member, IEEE*

***Abstract*—This paper proposes a Gaussian-Cauchy mixture maximum correntropy criterion Kalman filter algorithm (GCM_MCCKF) for robust state estimation in linear systems under non-Gaussian noise, particularly heavy-tailed noise. The performance of the MCCKF depends on the choice of kernel type and its associated kernel parameters. First, in terms of kernel type, it affects the sensitivity of the filter to noise and outliers. To overcome the limitations of a single zero-mean Gaussian kernel function in MCC, this paper proposes and derives the Gaussian-Cauchy mixture kernel function-based MCCKF. Some important properties of the Gaussian-Cauchy mixture correntropy are also studied. Second, in terms of kernel parameters, this paper draws on established techniques while incorporating innovative elements, heuristically proposes a suitable adaptive update scheme for kernel size to overcome the limitations of fixed kernel size in practical applications. Finally, the target tracking simulation example is used to verify that the proposed GCM_MCCKF algorithm can handle heterogeneous and complex data more flexibly and stably under the proposed kernel size adaptive update scheme, and obtain better filtering performance than the traditional KF filter, variational Bayesian filtering (VB), particle filter (PF), minimum error entropy KF (MEE_KF), single Gaussian kernel MCCKF (G_MCCKF) and double-Gaussian mixture MCCKF (DGM_MCCKF).**

***Index Terms*—Kalman filter, maximum correntropy criterion (MCC), Gaussian-Cauchy mixture kernel function, kernel size adaptive.**

Received 13 May 2024; revised 13 August 2024; accepted 5 October 2024. Date of publication 16 October 2024; date of current version 27 December 2024. This work was supported by the Qing Lan Project of Jiangsu Province under Grant R2023Q07, and the National Natural Science Foundation of China under Grant 62033010. The associate editor coordinating the review of this article and approving it for publication was Prof. Wei Yi. *(Corresponding author: Quanbo Ge.)*

Quanbo Ge is with the School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China, with Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing 210044, China, and also with the Jiangsu Provincial University Key Laboratory of Big Data Analysis and Intelligent Systems, Nanjing 210044, China (e-mail: quanboge@163.com).

Xuefei Bai is with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China, and also with the School of Electrical Engineering, Yancheng Institute of Technology, Yancheng 224051, China (e-mail: baixuefei005@163.com).

Pingliang Zeng is with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: plzeng@hotmail.com).

Digital Object Identifier 10.1109/TSP.2024.3479723

I. Introduction

FAST, accurate, and robust dynamic state estimators (DSEs) play an important role in many practical applications such as system control, target tracking, navigation, signal processing, information fusion, fault diagnosis, etc. [1], [2], [3], [4], [5], [6]. The Kalman filter, based on linear systems and Gaussian noise assumptions, stands as a landmark in this field, providing efficient recursive estimation by minimizing mean square error (MSE). However, in real-world scenarios, noise often exhibits a non-Gaussian distribution due to factors such as unreliable sensors, measurement errors, and impulsive noise. The heavy-tailed distribution, as a typical non-Gaussian distribution, not only includes impulsive noise as a special form of heavy-tailed noise but also may manifest as frequent moderate outliers. Its slowly decaying tails and numerous outliers pose challenges to traditional filtering methods based on mean squared error (MSE), particularly because these methods are sensitive to outliers. Higher-order statistics beyond mean and variance are necessary for characterizing heavy-tailed distributions. To address this, researchers have explored information theory learning (ITL) criteria such as the maximum correntropy criterion (MCC) and minimum error entropy (MEE) criterion in heavy-tailed non-Gaussian filtering [7], [8], [9], [10]. While the MEE criterion involves computationally expensive double summation operations, the maximum correntropy criterion (MCC) has garnered more attention for its effectiveness by extracting information from Gaussian deviations, thus offering improved performance over MSE-based approaches [11].

In previous work, the C-Filter was proposed for linear non-Gaussian systems, but its implementation overlooked the propagation of the estimation error covariance matrix [12], [13]. This oversight could degrade estimation accuracy [14], [15]. To rectify this limitation and account for error covariance, a fixed-point iteration technique was employed to develop maximum correntropy Kalman filters (MCKFs), which exhibited satisfactory estimation performance. Subsequently, MCKF was extended to nonlinear non-Gaussian systems [16]. In recent years, several MCKF variants have emerged. For instance, the maximum correntropy extended Kalman filter (MCEKF) [17], [18], maximum correntropy unscented Kalman filter (MCUKF) [19], [20], [21], maximum correntropy cubature Kalman filter

(MCCKF) [22], and maximum square root cubature Kalman filter (MCSCKF) [23] have all demonstrated effectiveness in nonlinear non-Gaussian scenarios. However, their performance may deteriorate in more complex noisy environments, such as those involving bimodal non-Gaussian noise.

In the field of maximum correntropy criterion (MCC) Kalman filtering, the kernel function assumes a pivotal role in gaugeing the resemblance between two data points, thereby influencing the filter estimate update and the impact of observed data on current state estimation [24]. The choice of kernel function holds paramount importance for MCC. Although the single Gaussian kernel function is frequently adopted due to its smoothness and strict positive determinism, it exhibits certain limitations.

1) Gaussian kernel functions are highly sensitive to kernel size, influencing MCC performance surfaces, the presence of local optima, convergence speed, and robustness against pulse noise [25]. Consequently, selecting an appropriate kernel size is imperative, yet definitive guidelines for this selection are currently lacking. Hence, it is crucial to opt for kernel functions relatively impervious to kernel size variations.
2) Single Gaussian kernel types may falter when handling complex data, potentially leading to performance degradation.
3) The impact of kernel size on filtering performance and convergence speed is contradictory within the effective data range, posing challenges for single Gaussian kernel functions in achieving a balance between performance and speed.

To overcome the limitations of single Gaussian kernel functions and improve filtering performance and adaptability, highlight the potential of kernel-based correntropy methods in enhancing the robustness of state estimation against non-Gaussian noise, researchers have introduced various alternative kernel functions to reduce sensitivity to kernel size. For example, the exponential kernel function is defined as $E_\sigma(\mathbf{e}) = \exp\left(-\frac{\|\mathbf{e}\|}{2\sigma^2}\right)$, where the L_1 norm replaces the L_2 norm to reduce dependence on kernel size parameters. However, its applicability is relatively limited. To further reduce sensitivity to kernel size, the exponential kernel function has been replaced by the Laplace kernel function [26], [27], $L_\sigma(\mathbf{e}) = \exp\left(-\frac{\|\mathbf{e}\|}{\sigma}\right)$, but its computational complexity is high when dealing with large-scale datasets. Additionally, when $x = y$, the derivative of the Laplace kernel function does not exist, leading to instability in optimization algorithms such as gradient descent. This instability can lead to convergence issues or cause the algorithm to get stuck in local optima, especially when using gradient-based optimization methods. The generalized Gaussian density (GGD) function proposed in the references [28], [29] demonstrates significant advantages when used as a kernel function for handling non-Gaussian noise and outliers, but it also brings challenges such as complex parameter selection and high computational complexity. The reference [30] proposes a maximum correntropy Kalman filter based on the Student's t-kernel to improve state estimation accuracy in the presence of outliers and non-Gaussian noise, but it also faces issues such as parameter selection, computational complexity, and non-convex optimization problems. To address the problem of singular matrices caused by multidimensional non-Gaussian noise, a new type of maximum correntropy Kalman filter based on the Cauchy kernel was proposed to enhance filter stability [31]. The Cauchy kernel function, due to its heavy-tailed characteristics, performs well in handling heavy-tailed noise distributions and provides better robustness. Reference [32] proposed a robust multi-innovation identification method based on Cauchy kernel correntropy to improve parameter estimation accuracy of nonlinear autoregressive models under non-Gaussian noise conditions. Additionally, reference [33] uses an elastic dynamic state estimation method based on the maximum correntropy cubature Kalman filter with a Cauchy kernel, showing better performance in power systems under non-Gaussian noise and denial-of-service attacks. On the other hand, researchers have introduced the concept of mixture correntropy (MC), where a convex combination of multiple kernel functions (homogeneous or heterogeneous) is employed within MCC, providing a more flexible and robust approach to handle various types of noise and outliers. This approach differs significantly from traditional single Gaussian kernel functions. Chen et al. in reference [34] first proposed a mixture correntropy function based on the convex combination of homogeneous double-Gaussian mixture kernel with different kernel sizes, pioneering the field of mixture correntropy filtering. References [35], [36], [37] also proposed double-Gaussian mixture correntropy for state estimation or machine learning. However, the homogeneous double-Gaussian mixture kernel functions in [34], [35], [36], [37] are essentially based on a single type of Gaussian kernel, achieving better robustness than a single kernel using the attributes of different kernel parameters. When dealing with complex data, especially data that are insensitive to kernel parameters, there is a risk of performance degradation. To address this issue, reference [38] utilized the Gaussian-Laplace heterogeneous kernel function convex combination maximum correntropy SVM framework, not only improving the limitations of correntropy under a single kernel, but also handling heterogeneous data more flexibly and stably. In reference [39], the mixture correntropy criterion based on double-Gaussian and Gaussian-Laplace mixtures achieved good results in robust filtering for nonlinear state-space models with measurement outliers. However, due to the inclusion of the Laplace kernel in [38] and [39], whose derivative does not exist at zero, gradient-based methods cannot be used to solve these problems, the semi-quadratic optimization algorithm, known for its rapid convergence, is employed [40], [41]. In reference [29], a maximum correntropy criterion based on the convex combination of generalized Gaussian kernels was proposed to improve active control of impulse noise, achieving fast convergence and low steady-state error. However, this method involves numerous parameters and high computational complexity, requiring effective optimization strategies to balance performance and real-time capability.

In the field of kernel size selection within the kernel function, the MCC algorithm's performance is significantly influenced. However, current literature and engineering practices often employ fixed-size kernels, selected empirically or via trial-and-error methods tailored to specific non-Gaussian noise types. Nevertheless, noise characteristics are inherently volatile, with initial high noise levels tending to stabilize over time. Consequently, the use of fixed-size kernels may not yield optimal performance. Moreover, fixed kernel sizes introduce a trade-off between convergence speed and steady-state bias [42].

- To better capture instantaneous error signals, time-varying kernel width (TKW) methods have been proposed. For instance, the variable kernel width MCC (VKW_MCC) method, introduced in [42], dynamically adjusts kernel size as a function of error, maximizing the Gaussian kernel function's influence for optimal kernel width.
- Adaptive kernel size methods, such as adaptive kernel size MCC (AMCC) [43], dynamically compute kernel size as a function of instantaneous error, aiming to strike a balance between performance and robustness. Similarly, the switch kernel size MCC (SMCC) method, proposed in [44], adaptively selects kernel size based on predicted kernel size and instantaneous error.

However, these adaptive kernel size methods may exhibit poor robustness or yield large steady-state errors under strong pulse conditions.

Given the challenges associated with kernel types and parameters, it is crucial to develop algorithms within the MCC framework that blend multiple kernel functions. These algorithms must be robust to kernel parameter variations and capable of handling complex datasets. Currently, there is a lack of MCC applications employing Gaussian-Cauchy mixture kernel functions for state estimation in target tracking amid non-Gaussian noise environments. To improve the performance and flexibility of the state estimation, this paper proposes a Kalman filter algorithm based on MCC with Gaussian-Cauchy mixture kernel functions. The algorithm selects a weighted sum of mixture kernel functions using prediction errors and residuals as a cost function for measurement updates in linear non-Gaussian systems, along with heuristically determined kernel size adaptation rules. Simulation results under various non-Gaussian noise scenarios, and the simulation results of manoeuvring target tracking show the effectiveness of the proposed GCM_MCCKF.

The remainder of this paper is organized as follows. Section II briefly reviews correntropy and introduces Gaussian-Cauchy mixture correntropy along with its properties. Section III proposes and derives an MCC Kalman filtering algorithm based on Gaussian-Cauchy mixture kernel functions (GCM_MCCKF), and incorporating existing techniques with innovative elements to design an adaptive kernel size mechanism. Section IV validates the effectiveness of the algorithm under the adaptive kernel size update mechanism through simulations, further demonstrating its superior robustness in complex noise environments. Finally, Section V concludes the paper and outlines future research directions.

II. Mixture Correntropy

A. Correntropy

Given two random variables $\mathbf{X} \in \mathbb{R}$ and $\mathbf{Y} \in \mathbb{R}$, the correntropy [45] is defined as

$$V(\mathbf{X},\mathbf{Y}) = E[\kappa(\mathbf{X},\mathbf{Y})] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \kappa(x,y)p(x,y)dxdy \tag{1}$$

where E[·] denotes the expectation operation and $p(x,y)$ is the joint probability density function (PDF) of $\mathbf{X}$ and $\mathbf{Y}$. $\kappa(x,y) = \kappa(x-y)$ is a shift-invariant Mercer kernel function, usually defaulted to a Gaussian kernel function $G_\sigma(x-y)$ with a kernel size of σ.

$$\kappa_\sigma(x,y) = G_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right) \tag{2}$$

where $e = x - y$ is the error, $\|\cdot\|$ denotes the 2-norm. It is evident that with increasing the error variable e, the Gaussian kernel function gradually converges to zero, rendering it less sensitive to outliers. Gaussian kernel function-based correntropy $V(\mathbf{X},\mathbf{Y})$ is often defined as the Gaussian correntropy, represented by $G(\mathbf{X},\mathbf{Y})$.

In practice, the joint PDF of the data is unknown in advance, and only a limited number of data samples are available. In this case, the Gaussian correntropy needs to be estimated using the sample mean based on the Parzen window method. The estimation can be expressed as follows:

$$\hat{G}_\sigma(\mathbf{X},\mathbf{Y}) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e_i) \tag{3}$$

where $e_i = x_i - y_i$ represents the sample errors, and $\{\mathbf{x}_i,\mathbf{y}_i\}_{i=1}^N$ are the N data samples drawn from $p(x,y)$. Further discussions on the properties of Gaussian correntropy and MCC are detailed in [24].

From (3), it's evident that correntropy outperforms MSE in the presence of outliers. Error samples are weighted by a Gaussian kernel, with smaller errors contributing significantly to the objective function. Large errors or outliers are given smaller weights and filtered out by the Gaussian kernel. Outliers outside the kernel's neighborhood are assigned zero weight, hence don't affect the objective function. Correntropy, with its selective weighting mechanism, is less sensitive to outliers compared to the uniform weighting of MSE, making it more robust against non-Gaussian noise and less likely to degrade in performance when confronted with significant outliers.

By taking the Taylor series expansion of the Gaussian correntropy, we obtain:

$$G_\sigma(\mathbf{X},\mathbf{Y}) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n\sigma^{2n}n!} E\left[(\mathbf{X}-\mathbf{Y})^{2n}\right] \tag{4}$$

From (4), it's evident that the Gaussian correntropy incorporates even-order moments of the error variable $(\mathbf{X}-\mathbf{Y})$, weighted by kernel size σ. Choosing an appropriate σ

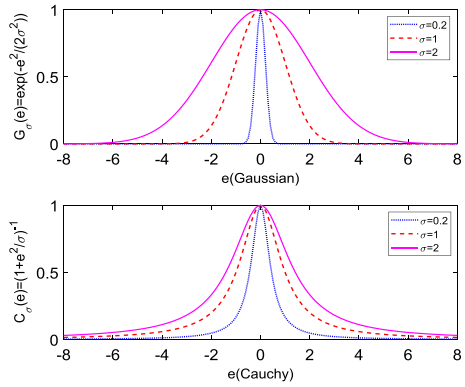

Fig. 1. Comparison of Gaussian and Cauchy kernel functions.

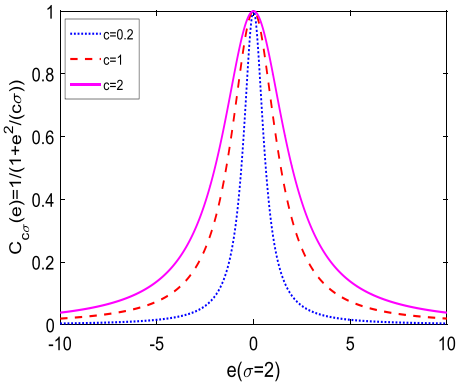

Fig. 2. The change of Cauchy kernel function with c.

enables capturing high-order moments of error variables, making correntropy-based Kalman filters more robust in non-Gaussian noise scenarios. Unlike traditional Gaussian-based filters, correntropy criteria can handle higher-order information, crucial in non-Gaussian scenarios.

The heavy-tailed distribution characteristics of non-Gaussian noise can significantly impact filtering performance, including accuracy (extreme values or outliers can introduce bias into filters, reducing signal estimation accuracy), stability (extreme values or outliers can increase signal volatility, leading to more unstable filter output), convergence rate (slow signal changes in heavy-tailed distributions require filters more time to adapt, affecting the speed of convergence and the performance in tracking signal variations) and robustness of estimation (outliers in heavy-tailed distributions can disrupt filter output, reducing the filter robustness). Therefore, it is crucial to carefully consider the distribution characteristics of the data in filter design, especially in the presence of heavy-tailed distributions, and appropriate methods should be employed to improve filter performance. Within the framework of MCC-based filters, the selection of suitable kernel functions should be considered to improve filter performance.

We are investigating the Cauchy kernel function, which excels in handling strong outliers or non-Gaussian noise. The Cauchy kernel function is defined as $C_\sigma(e) = \left(1 + \frac{e^2}{\sigma}\right)^{-1}$. Fig. 1 compares the Gaussian kernel $G_\sigma(e)$ and the Cauchy kernel $C_\sigma(e)$ in terms of their sensitivity to the kernel size parameter σ. As shown in Fig. 1, the Cauchy kernel function is less sensitive to the kernel size parameter compared to the Gaussian kernel and exhibits heavy-tailed characteristics. This heavy-tailed nature allows the Cauchy kernel function to effectively reduce the impact of outliers in the data, thereby improving filtering performance in the presence of strong outliers or non-Gaussian noise.

By studying the convexity of these two kernel functions, it can be seen that the Gaussian kernel is convex at $|e| \leq \sigma$, while the Cauchy kernel is convex at $|e| \leq \sqrt{\frac{\sigma}{3}}$, and their extreme values can be obtained by taking derivatives.

The heaviness of the tails in a Cauchy kernel function can be controlled by adjusting the heavy-tailed factor c. The Cauchy kernel function with this factor is represented as:

$$C_{c\sigma}(e) = \left(1 + \frac{e^2}{c\sigma}\right)^{-1} \tag{5}$$

The impact of varying the heavy-tailed factor c on the Cauchy kernel function is depicted in Fig. 2.

It is evident from the figure that increasing c results in thicker tails in the Cauchy kernel function.

Similar to the Gaussian kernel function, a Taylor series expansion is performed for the Cauchy kernel function, yielding:

$$C_{c\sigma}(\mathbf{X}, \mathbf{Y}) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(c\sigma)^n} E\left[(\mathbf{X} - \mathbf{Y})^{2n}\right] \tag{6}$$

It can be observed that Cauchy correntropy can be seen as a weighted sum of all even-order moments of $\mathbf{X} - \mathbf{Y}$, enabling it to extract higher-order statistics from the data.

B. Mixture Correntropy (MC)

Considering the limitations of single Gaussian kernel functions outlined in the introduction, some researchers have proposed using a combination of kernel functions in the correntropy criterion to improve the performance and flexibility of filtering in non-Gaussian noise scenarios. There are typically two forms of mixture correntropy:

- Homogeneous mixture kernel functions refer to the combination of the same type of kernel functions with different parameters (e.g., different kernel sizes) to form a mixed kernel function. This approach retains the characteristics of the original kernel functions while enhancing performance and adaptability by adjusting parameters. These mixture kernels have also been proposed for robust filtering and smoothing in nonlinear non-Gaussian systems.
- Heterogeneous mixture kernel functions refer to the combination of different types of kernel functions to form a mixed kernel function. This combination can include a convex mixture of various kernel functions (e.g., Gaussian and Laplacian kernels). The goal is to leverage the strengths of different kernel functions to enhance data

processing capabilities, particularly when dealing with complex, non-uniform data. By integrating different kernel functions, heterogeneous mixture kernel functions can improve the robustness and adaptability of algorithms, addressing issues that a single kernel function may not effectively handle.

By combining different types of kernel functions, the mixture correntropy method can leverage the strengths of each, improving the filter's adaptability and robustness in complex data environments. This approach also extends the application range of MCCKF algorithms, providing more reliable solutions for data processing in complex environments.

The general form of mixture correntropy, based on a convex combination of kernel functions, is formulated as follows:

$$M(\mathbf{X},\mathbf{Y}) = E\left(\sum_{i=1}^{N}\alpha_i\kappa_i(\mathbf{X},\mathbf{Y})\right) \tag{7}$$

where $N \geq 2$ is the number of kernel functions, $\alpha_i \geq 0$ with $\sum_{i=1}^{N}\alpha_i = 1$ are the mixing coefficients, and $\kappa_i(\mathbf{X},\mathbf{Y})$ denotes the i-th kernel function. Although multiple kernel functions can be combined, it increases the number of free parameters and consequently complexity. Hence, for simplicity, only the combination of two kernel functions is discussed herein.

Typically, mixture correntropy composed of two kernel functions manifests itself in two forms:

1) $\kappa_1(\cdot)$ and $\kappa_2(\cdot)$ are combinations of the same type of kernel functions with different kernel sizes, known as homogeneous mixture correntropy.
2) $\kappa_1(\cdot)$ and $\kappa_2(\cdot)$ are combinations of different types of kernel function, known as heterogeneous mixture correntropy.

The sample-based mixture correntropy kernel function used in practical applications can be expressed as

$$M[\mathbf{X},\mathbf{Y}] = \frac{1}{N}\sum_{i=1}^{N}[\alpha\kappa_1(e) + (1-\alpha)\kappa_2(e)] \tag{8}$$

where $e = \mathbf{x} - \mathbf{y}$. When $\alpha = 0$ or $\alpha = 1$, the mixture kernel function degenerates into the corresponding original kernel function.

C. Gaussian-Cauchy Mixture Correntropy

Based on the issues identified with the existing mixture kernel functions and the preliminary research on Cauchy kernel functions, this paper proposes the exploration of a Gaussian-Cauchy heterogeneous mixed correntropy Kalman filter (GCM_MCCKF) for robust state estimation using gradient-based methods in non-Gaussian linear systems to enhance performance. Currently, there are few cases of applying Gaussian-Cauchy mixed kernel functions in mixed correntropy to enhance the robustness of state estimation under heavy-tailed non-Gaussian noise, thus making the study of GCM_MCCKF highly significant.

The Gaussian-Cauchy mixture correntropy is formulated as follows:

$$M_{GCM}(e) = E\left[\alpha G_\sigma(e) + (1-\alpha)C_{c\sigma}(e)\right] \tag{9}$$

The sample estimate for the Gaussian-Cauchy mixture correntropy is

$$\hat{M}_{GCM}(e) = \frac{1}{N}\sum_{i=1}^{N}\left[\alpha\exp\left(-\frac{e^2}{2\sigma^2}\right) + (1-\alpha)\frac{1}{1+\frac{e^2}{c\sigma}}\right] \tag{10}$$

Gaussian-Cauchy mixture kernel function combines the smoothness and mathematical tractability of the Gaussian kernel with the heavy-tailed characteristics against outliers of the Cauchy kernel, providing a more flexible and robust approach to handle various types of noise and outliers. Specifically, this approach is designed to:

1) Provide strong performance in the presence of primarily Gaussian noise using the Gaussian component, while mitigating the impact of heavy-tailed outliers in the data with the Cauchy component;
2) Leverage the heavy-tailed nature of the Cauchy kernel to enhance the filter's tolerance towards outliers, thus providing more stable and accurate estimations in complex noise environments;
3) Adjust the ratio or parameters of the Gaussian and Cauchy kernels to maintain optimal performance in different noise scenarios.

D. Properties of Gaussian-Cauchy Mixture Correntropy

In this subsection, we present some properties of the proposed Gaussian-Cauchy Mixture Correntropy.

Property 1: When $\alpha = 1$, the Gaussian-Cauchy mixture kernel reduces to a single Gaussian kernel. For $\alpha = 0$, it corresponds to the Cauchy kernel with a kernel size σ and a heavy-tailed factor c.

Property 2: The Gaussian-Cauchy mixture kernel is less sensitive to kernel size compared to the Double-Gaussian mixture kernel, exhibiting more pronounced heavy-tailed characteristics.

Property 3: Near $e = 0$, the Gaussian-Cauchy mixture kernel demonstrates both smoothness and convexity.

Property 4: Symmetry-$M_{GCM}[\mathbf{X},\mathbf{Y}]$ is symmetric, i.e., $M_{GCM}[\mathbf{X},\mathbf{Y}] = M_{GCM}[\mathbf{Y},\mathbf{X}]$.

Property 5: Positivity and Boundedness - $M_{GCM}[\mathbf{X},\mathbf{Y}]$ is positive and bounded. It reaches its maximum $M_{GCM}[\mathbf{X},\mathbf{Y}] = 1$ if and only if $\mathbf{X} = \mathbf{Y}$. When $\mathbf{X}$ and $\mathbf{Y}$ are sufficiently apart, $M_{GCM}[\mathbf{X},\mathbf{Y}]$ tends to but cannot reach 0.

Property 6: Translation Invariance - $M_{GCM}[\mathbf{X},\mathbf{Y}]$ has translation invariance, i.e., $M_{GCM}[\mathbf{X}-\mathbf{Z},\mathbf{Y}-\mathbf{Z}] = M_{GCM}[\mathbf{X},\mathbf{Y}]$.

Property 7: Even Moments - $M_{GCM}[\mathbf{X},\mathbf{Y}]$ involves all even moments of error e, as given by Equation (11).

$$M_{GCM}(e) = \sum_{n=0}^{\infty}\frac{(-1)^n\left[\alpha c^n + (1-\alpha)(2c)^n n!\right]}{(2c)^n\sigma^{2n}n!}E\left[e^{2n}\right] \tag{11}$$

Property 8: Approximation - When σ is large enough (i.e., e is small enough), the approximation in Equation (12) holds.

$$M_{GCM}[\mathbf{X},\mathbf{Y}] \approx 1 - \left(\frac{\alpha}{2\sigma^2} + \frac{1-\alpha}{c\sigma}\right)e^2 \tag{12}$$

Proof: By approximating $\exp(x) \approx 1 + x$ and $(1+x)^n \approx 1 + nx$ for small x, we obtain:

$$\begin{aligned} M_{GCM}[\mathbf{X},\mathbf{Y}] &= \alpha G_\sigma(e) + (1-\alpha)C_{c\sigma}(e) \\ &\approx \alpha\left(1 - \frac{e^2}{2\sigma^2}\right) + (1-\alpha)\left(1 - \frac{e^2}{c\sigma}\right) \\ &= 1 - \left(\frac{\alpha}{2\sigma^2} + \frac{1-\alpha}{c\sigma}\right)e^2 \end{aligned} \tag{13}$$

This completes the proof, indicating that when the kernel size is sufficiently large, MCC is equivalent to MSE.

Property 9: Convexity - As a function of e, $M_{GCM}[\mathbf{X},\mathbf{Y}]$ is convex at any point satisfying $|e| \le \sigma$ (when $c \ge 3\sigma$) and $|e| \le \sqrt{\frac{c\sigma}{3}}$ (when $c \le 3\sigma$).

Proof: The Hessian matrix of $M_{GCM}[\mathbf{X},\mathbf{Y}] = M_{GCM}[e]$ with respect to e is given by:

$$\begin{aligned} H_{M_{GCM}[e]} &= \left[\frac{\partial^2 M_{GCM}[e]}{\partial^2 e}\right] \\ &= -\left[\alpha\left(\sigma^2 - e^2\right)\frac{G_\sigma(e)}{\sigma^4}\right. \\ &\quad \left. + (1-\alpha)\left(c\sigma - 3e^2\right)\frac{2c\sigma}{\left(c\sigma + e^2\right)^3}\right] \\ &= -\left[\alpha\left(\sigma^2 - e^2\right)\xi + (1-\alpha)\left(c\sigma - 3e^2\right)\zeta\right] \end{aligned} \tag{14}$$

where $e = \mathbf{X} - \mathbf{Y}$, $\xi = \frac{G_\sigma(e)}{\sigma^4} > 0$, and $\zeta = \frac{2c\sigma}{(c\sigma + e^2)^3} > 0$. When $H_{M_{GCM}[e]} \le 0$, i.e., when $\alpha(\sigma^2 - e^2)\xi + (1-\alpha)(c\sigma - 3e^2)\zeta \ge 0$, $M_{GCM}[e]$ is a convex function.

As mentioned above, the Gaussian kernel function is convex for $|e| \le \sigma$, while the Cauchy kernel function is convex for $|e| \le \sqrt{\frac{c\sigma}{3}}$. Thus, when $c \ge 3\sigma$, we have $\sqrt{\frac{c\sigma}{3}} \ge \sigma$, which implies that the Gaussian-Cauchy mixture kernel is convex for any point with $|e| \le \sigma$. Similarly, when $c \le 3\sigma$, we have $\sqrt{\frac{c\sigma}{3}} \le \sigma$, indicating that the Gaussian-Cauchy mixture kernel is convex for any point with $|e| \le \sqrt{\frac{c\sigma}{3}}$.

This completes the proof.

III. Gaussian-Cauchy Mixture Based Maximum Correntropy Kalman Filtering

In this section, the discrete-time dynamic model of the linear system is considered and the MCC-KF algorithm based on the Gaussian-Cauchy mixture kernel is derived.

A. Linear Discrete-Time State-Space Model

Consider the following linear dynamical system model with non-Gaussian noise.

$$\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \boldsymbol{\omega}_{k-1} \tag{15}$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\upsilon}_k \tag{16}$$

where $\mathbf{y}_k \in \mathbb{R}^m$ denotes a measurement vector related to the state vector of interest $\mathbf{x}_k \in \mathbb{R}^n$ at time k. $\mathbf{F}_k \in \mathbb{R}^{n\times n}$ and $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ denote the state transition matrix and the observation matrix (measurement matrix), respectively. $\boldsymbol{\omega}_{k-1} \in \mathbb{R}^n$ and $\boldsymbol{\upsilon}_k \in \mathbb{R}^m$ denote mutually uncorrelated additive non-Gaussian process and measurement noises, which have a nominal process noise covariance matrix $\mathbf{Q}_k \in \mathbb{R}^{n\times n}$ and a nominal measurement noise covariance matrix $\mathbf{R}_k \in \mathbb{R}^{m\times m}$, respectively. These matrices represent the expected noise characteristics under ideal conditions, assuming that the noise follows a Gaussian distribution. However, in practical scenarios, noise often deviates from Gaussianity, exhibiting heavy-tailed or skewed characteristics. Non-Gaussian heavy-tailed noise is typically represented using the following Gaussian mixture model (GMM), allowing for more accurate modeling of the real-world noise behavior.

$$\boldsymbol{\omega}_k \sim \sum_{i=1}^{M}\pi_i\left((1-a_i)\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\omega}_i}, \mathbf{Q}_k) + a_i\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\omega}_{i,\text{ heavy}}}, \Phi_i\mathbf{Q}_k)\right) \tag{17}$$

$$\boldsymbol{\upsilon}_k \sim \sum_{j=1}^{N}\pi_j\left((1-b_j)\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\upsilon}_j}, \mathbf{R}_k) + b_j\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\upsilon}_{j,\text{heavy}}}, \Omega_j\mathbf{R}_k)\right) \tag{18}$$

where $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ represents a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ. The coefficients a_i and b_j, where $a_i, b_j \in [0,1]$, denote the Gaussian pollution level coefficients for the i-th and j-th components, used to adjust the degree of Gaussian pollution. $\boldsymbol{\mu}_{\boldsymbol{\omega}_{i,\text{heavy}}}$ and $\boldsymbol{\mu}_{\boldsymbol{\upsilon}_{j,\text{heavy}}}$ are the means of the heavy-tailed distributions for the i-th and j-th components, respectively. Φ_i and Ω_j are the corresponding covariance multipliers. M and N denote the number of Gaussian components, respectively, and π_i and π_j represent the weights of each component, satisfying $\sum_{i=1}^{M}\pi_i = 1$ and $\sum_{j=1}^{N}\pi_j = 1$.

Nominal noise covariance is typically used under the assumption that the noise follows a Gaussian distribution. However, In practical scenarios, noise often exhibits non-Gaussian characteristics, such as heavy tails or the presence of outliers. Consequently, noise covariance is usually unknown, time-varying or corrupted by outliers due to complex environments or measurement equipment nonidealities (aging, lack of calibration, manufacturing issues, error accumulation, failure, etc.), operator errors, environmental interference, etc. These factors can significantly degrade estimation accuracy. Determining the matrix parameters of heavy-tailed noise remains an open issue. As documented in [46], [47] and [48], the variational Bayesian (VB) approach approximates the covariance of unknown and time-varying noise by modeling the filtering distribution as a product of Gaussian and inverse Wishart (IW) distributions, a method that has been proven to be efficient. However, in cases where non-Gaussian noise contains significant outliers or shot

noise, the accuracy of estimation using the VB approximation method will be compromised. Numerous studies have resorted to simplify the entire Gaussian Mixture Model (GMM) to a single Gaussian, potentially leading to the loss of crucial information inherent in the mixture components. In this paper, we use an improved robust Expectation Maximization (EM) algorithm in our previous work [49] to estimate the parameters of a Gaussian mixture mode. This improved algorithm overcomes the limitations of the traditional EM algorithm, such as the need to prespecify the specific number of mixing components, the sensitivity to initial parameters, and the susceptibility to being stuck in local optima. "∧" above a parameter symbol to denote the parameter estimated by the improved EM algorithm: for example, the estimated covariance of the process noise $\mathbf{Q}_k$ is represented as $\hat{\mathbf{Q}}_k$. Subsequently, a fusion approach is employed for the combination of Gaussian terms, which is based on the Gaussian term merging method that utilizes both the Mahalanobis distance and the Kullback-Leibler (KL) distance. (For a detailed process, see [49]).

B. Gaussian-Cauchy Mixture Based MCC-KF Algorithm

In this subsection, we will present the MCC-KF algorithm that uses the Gaussian-Cauchy mixture kernel.

Compared to traditional Kalman filtering, the MCC-based Kalman filtering algorithm is the same in the prediction stage and different in the update stage. The GCM_MCCKF estimation method, reliant on mixture kernels, employs Gaussian-Cauchy mixture kernel functions of the error signal vector instead of the quadratic function.

1) Prediction stage (Time update stage)

The purpose is to calculate a prior estimate $\hat{\mathbf{x}}_k^-$ (predicting state variables) and covariance $\mathbf{P}_{k|k-1}$ (predicting covariance).

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_k \hat{\mathbf{x}}_{k-1} \tag{19}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \hat{\mathbf{Q}}_k \tag{20}$$

where $\hat{\mathbf{x}}_k^-$ and $\hat{\mathbf{x}}_{k-1}$ are the one-step predictions of the state at time k and the estimated state variable at time $k-1$, respectively. $\mathbf{P}_{k|k-1}$ is the prediction covariance of the error at time k and $\mathbf{P}_{k-1}$ is the covariance of the state estimation error at time $k-1$.

2) Measurement Update Stage

The purpose is to calculate a posterior state estimate $\hat{\mathbf{x}}_k$ and covariance $\mathbf{P}_k$.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_{GCM,k}(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-) \tag{21}$$

$$\mathbf{K}_{GCM,k} = \mathbf{P}_{k|k-1} \Psi_{GCM,k} \mathbf{H}_k^T (\hat{\mathbf{R}}_k + \mathbf{H}_k \mathbf{P}_{k|k-1} \Psi_{GCM,k} \mathbf{H}_k^T)^{-1} \tag{22}$$

where $\Psi_{GCM,k} = \dfrac{A_k}{B_k}$, $A_k = \dfrac{\alpha G_{\sigma\_r}}{\sigma^2} + \dfrac{2(1-\alpha) C_{c\sigma\_r}^2}{c\sigma}$ and $B_k = \dfrac{\alpha}{\sigma^2} + \dfrac{2(1-\alpha)}{c\sigma}$.

$$\mathbf{P}_k = (\mathbf{I}_n - \mathbf{K}_{GCM,k}\mathbf{H}_k)\, \mathbf{P}_{k|k-1}\, (\mathbf{I}_n - \mathbf{K}_{GCM,k}\mathbf{H}_k)^T + \mathbf{K}_{GCM,k} \hat{\mathbf{R}}_k \mathbf{K}_{GCM,k}^T \tag{23}$$

The detailed derivation process of (21), (22) and (23) are shown in Appendix A. Equations (21), (22), and (23) are equivalent to Equations (28), (30) and (33) in Appendix A, respectively.

C. Select Adaptive Kernel Size

As an important parameter in the mixture function-based MCC filtering algorithms, the kernel size significantly impacts filter performance [50], [51]. The limitations of using a fixed kernel size have been discussed previously. Our method for selecting the adaptive kernel size draws on established techniques while incorporating innovative elements to improve filtering performance.

As described in [13], the adaptive kernel size is determined based on the Euclidean distance $\sigma_k = \|\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k\|$, offering a simple yet effective approach considering available measurements and the system model perspective. Based on this, reference [52] further refines the adaptive kernel size using the Mahalanobis distance $\sigma_k = \|\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k\|_{\mathbf{R}_k^{-1}}$, which incorporates noise parameters into the adaptation process, leading to improved filtering accuracy.

In the Maximum Correntropy Criterion-based Kalman filter, correntropy acts as a cost function that affects the filter gain $\mathbf{K}$ during the measurement update stage. A comparison of [13] and [52] reveals that the critical parameter $\mathbf{L}_k$, which affects $\mathbf{K}$, maintains consistency $\left(\mathbf{L}_k = G_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right) = \exp\left(-\frac{1}{2}\right)\right)$ under two different adaptive kernel size rules, as follows:

$$\mathrm{L}_{k,\mathrm{Euclidean}} = \exp\left(-\frac{\|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|^2}{2\|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|^2}\right)$$

$$\mathrm{L}_{k,\mathrm{Mahalanobis}} = \exp\left(-\frac{\|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|_{\mathbf{R}_k^{-1}}^2}{2\|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k\|_{\mathbf{R}_k^{-1}}^2}\right)$$

Further analysis of the exponential function $\exp(-x)$ underlying these adaptations reveals that higher values of x (where $x > 0$) lead to a decrease in the function value, thus reducing $\mathbf{L}_k$ and consequently the filter gain, directly improving filter performance. To further advance this method, our approach incorporates the Mahalanobis distance along with a mixture coefficient α and a heavy-tailed factor c in the Gaussian-Cauchy mixed kernel function, effectively increasing the value of x in the exponential function. Our preliminary simulations using this adapted criterion from [52] have shown promising results. Through these experiments, we also observed that adjusting the heavy-tail factor and the mixture coefficient within the mixed kernel function can affect filtering performance. Based on these findings, we have developed a new rule for the adaptive kernel size.

$$\sigma_{new,k} = \frac{\alpha}{c} \|\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k\|_{\hat{\mathbf{R}}_k^{-1}} \tag{24}$$

where α and c are the mixture coefficient and the heavy-tail factor in the Gaussian-Cauchy mixed kernel function, respectively. The term $\|\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k\|_{\hat{\mathbf{R}}_k^{-1}}$ denotes the Mahalanobis

distance of the residual vector $\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k$ with respect to the inverse covariance matrix $\hat{\mathbf{R}}_{k_{\hat{}}}^{-1}$, which means that when the residuals are large, the role of $\hat{\mathbf{R}}_k^{-1}$ is to assign smaller weights to these large residuals, thereby reducing the influence of outliers on the filtering results. In this way, the filter can more robustly handle non-Gaussian noise and outliers, improving the accuracy and stability of the filtering. On the other hand, by incorporating the estimated noise covariance $\hat{\mathbf{R}}_k$, which better reflects the non-Gaussian characteristics of noise in practical scenarios compared to the uncontaminated nominal noise covariance $\mathbf{R}_k$ based on Gaussian assumptions used in reference [52], we can dynamically adjust the filtering process to better handle non-Gaussian noise, especially when dealing with outliers or heavy-tailed distributions, thus improving filtering accuracy and robustness. Furthermore, the introduction of the mixing coefficient α and the heavy-tail factor c in the Gaussian-Cauchy mixture kernel function further optimizes the adaptation process, making our approach more effective in addressing non-Gaussian noise characteristics. Experimental results have shown that our adaptive kernel size method outperforms the traditional Mahalanobis distance rule used in reference [52]. This effectiveness arises from our method's ability to more efficiently reduce filter gain in high-noise scenarios, thereby enhancing overall filtering performance. Under this new rule, in the presence of heavy-tailed noise scenarios, with $c \geq 1$ and $\alpha \leq 1$, the parameter satisfies $L_{k-\text{new}} \leq L_{k-\text{Mahalanobis}}$, with $L_{k-\text{new}}$ being significantly smaller for larger values of c. $L_{k-\text{new}}$ corresponds to the parameter $\Psi_{GCM,k}$ in Equation (22) of our algorithm. $\Psi_{GCM,k}$, as a matrix for adjusting the filtering gain, its appropriate size depends on the noise characteristics of the system and the requirement for estimation response speed. In scenarios with high noise levels or significant system uncertainties, a smaller $\Psi_{GCM,k}$ may be more beneficial for maintaining the stability of the filtering performance. Conversely, in scenarios where the system is highly dynamic and noise levels are low, appropriately increasing $\Psi_{GCM,k}$ may enhance the real-time performance of the estimation. Therefore, the specific impact of making $\Psi_{GCM,k}$ larger or smaller on filtering performance needs to be balanced and adjusted based on the actual application context.

IV. Simulation Results

In this section, we are interested in the performance of the proposed kernel size adaptive GCM_MCCKF algorithm. Here, a dynamic linear tracking model is considered for land vehicles [14] (the known acceleration in the model $u_k = 0$). The simulation results in different noise scenarios are compared with traditional KF, variational Bayesian filtering [53], particle filter [54], minimum error entropy KF (MEE_KF) [10], single Gaussian kernel MCCKF (G_MCCKF) [14], and double-Gaussian mixture MCCKF (DGM_MCCKF) [34]. MATLAB 2018a simulations are executed on a PC equipped with an Intel(R) Core(TM) i5-8250U CPU running at 1.60GHz to 1.80GHz and having 12GB RAM.

A. Performance Metrics

The root mean squared error (RMSE) and the averaged RMSE (ARMSE) of the state are used as performance evaluation metrics.

$$RMSE_{\mathbf{x}_k} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\mathbf{x}_k^j - \hat{\mathbf{x}}_k^j)^2}$$

$$ARMSE_{\mathbf{x}_k} = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\mathbf{x}_k^j - \hat{\mathbf{x}}_k^j)^2}$$

where $\hat{\mathbf{x}}_k^j$ is the estimate of the true state variable $\mathbf{x}_k^j$ at time k in the j-th samples. M and N denote the number of Monte Carlo runs and the number of samples, respectively. We set $M = 100$ and $N = 200$ in this simulation.

B. Simulation Settings

Take the four-dimensional state variables as an example. The first two elements of the state variable are the north and east positions of a land vehicle, while the last two are the corresponding velocities. Therefore, $\mathbf{F}_k$ and $\mathbf{H}_k$ are represented as

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & \triangle t & 0 \\ 0 & 1 & 0 & \triangle t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

where $\triangle t = 2s$ is the sample time.

In this paper, based on the physical meaning of the system and practical application requirements, the initial state and covariance matrix are set as follows:

$$\begin{aligned} &\mathbf{x}_0 = \hat{\mathbf{x}}_0 = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^T \\ &\mathbf{P}_0 = \text{Diag}\left(\begin{bmatrix} 0.52 & 0.55 & 0.08 & 0.13 \end{bmatrix}\right) \\ &\mathbf{Q}_{k-1}^1 = \text{Diag}\left(\begin{bmatrix} 0.13 & 0.22 & 0.09 & 0.18 \end{bmatrix}\right) \\ &\mathbf{R}_k^1 = \text{Diag}\left(\begin{bmatrix} 0.8 & 2.1 \end{bmatrix}\right) \end{aligned}$$

where the initial state vector $\mathbf{x}_0$ represents a land vehicle starting from a known position (1,1) and at rest (velocity 0). This choice reflects a realistic scenario in which the vehicle starts from a known location. The initial error covariance matrix $\mathbf{P}_0$ indicates different levels of uncertainty for position and velocity, which is in line with practical considerations. The higher values (0.52 and 0.55) for the position components reflect greater initial uncertainty, while the lower values (0.08 and 0.13) for velocity components indicate more confidence in the initial velocity estimates. The process noise covariance matrix $\mathbf{Q}_{k-1}^1$ describes the random disturbances and uncertainties in the system process. For example, a value of 0.22 indicates a higher level of uncertainty for the corresponding state variable, while a value of 0.09 indicates lower uncertainty. This setting helps to accurately capture the dynamic changes in different state variables. The measurement noise covariance matrix $\mathbf{R}_k^1$ describes the noise levels in the measurement process. Larger

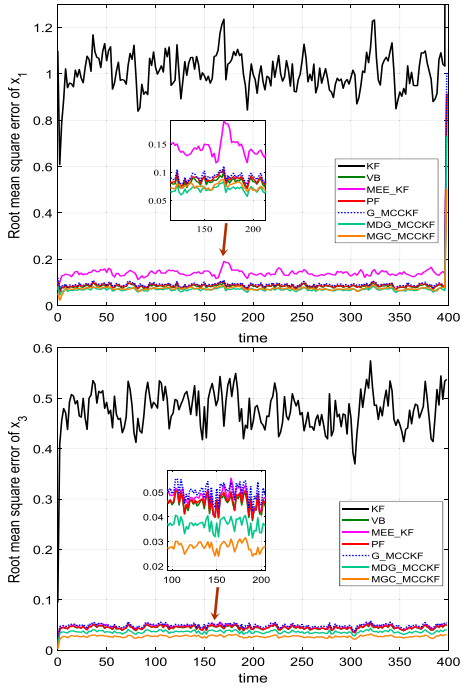

Fig. 3. RMSEs of state variables x_1 and x_3 under Gaussian noise.

values (2.1) indicate higher noise in that measurement dimension, while smaller values (0.8) indicate lower noise. This setting accurately reflects the noise characteristics of the measurement equipment or process, improving the interpretation and processing accuracy of the measurement data.

The proposed GCM_MCCKF algorithm employs the adaptive kernel size rule we introduced in (24): $\sigma_{\text{new,k}} = \frac{\alpha}{c} \|\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k\|_{\hat{\mathbf{R}}_k^{-1}}$. In contrast, the single Gaussian kernel MCCKF (G_MCCKF) and the double-Gaussian mixture MCCKF (DGM_MCCKF) algorithms utilize an adaptive kernel size rule based on the Mahalanobis distance: $\sigma_k = \|\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k\|_{\mathbf{R}_k^{-1}}$ with $\sigma_{2,k} = 20\sigma_{1,k}$ in DGM_MCCKF.

C. Simulation Results

A) To verify the superior performance of the proposed GCM-MCCKF algorithm in different noise scenarios.

1) Scenario 1: In Equations (15) and (16), the process noise $\boldsymbol{\omega}_{k-1}$ and the measurement noise $\boldsymbol{\upsilon}_k$ are both modeled as zero-mean Gaussian noise, described as:

$$\begin{cases} \boldsymbol{\omega}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1}^1) \\ \boldsymbol{\upsilon}_k \sim \mathcal{N}(0, \mathbf{R}_k^1) \end{cases}$$

where $\mathbf{Q}_{k-1}^1$ and $\mathbf{R}_k^1$ denote the covariance matrices of the process noise and measurement noise, respectively.

The RMSEs of the position state variable x_1 and the velocity state variable x_3 under seven filtering methods with Gaussian noise are shown in Fig. 3. Table I presents a comparison of the estimation performance of seven estimation algorithms under Gaussian noise. In this case, the covariance matrix in the algorithm derivation does not require estimation using the improved EM algorithm we adopted; the covariance for the Gaussian case can be used directly.

TABLE I
ARMSEs OF DIFFERENT FILTERS WITH GAUSSIAN NOISE

Algorithms	ARMSE			
	x_1	x_2	x_3	x_4
KF	1.0399	2.1633	0.47072	0.72725
VB	0.087719	0.12354	0.044462	0.062376
MEE_KF	0.14474	0.17999	0.045912	0.065013
PF	0.090733	0.1275	0.044797	0.062924
G_MCCKF	0.09649	0.1359	0.048908	0.068614
DGM_MCCKF	0.072714	0.10217	0.035852	0.050361
GCM_MCCKF	0.074886	0.1046	0.026823	0.037865

Table I and Fig. 3 show that the seven algorithms exhibit effective filtering performance under Gaussian noise, with the Kalman filter displaying slightly lower performance compared to the others, which show slightly better performance with a minor performance discrepancy.

2) Scenario 2: The process noise $\boldsymbol{\omega}_{k-1}$ and measurement noise $\boldsymbol{\upsilon}_k$ in Equations (15) and (16) as heavy-tailed non-Gaussian noises, originating from a zero-mean Gaussian mixture model

$$\begin{cases} \boldsymbol{\omega}_{k-1} \sim (1-a)\mathcal{N}\left(0, \mathbf{Q}_{k-1}^1\right) + a\mathcal{N}\left(0, \mathbf{Q}_{k-1}^2\right) \\ \boldsymbol{\upsilon}_k \sim (1-b)\mathcal{N}\left(0, \mathbf{R}_k^1\right) + b\mathcal{N}\left(0, \mathbf{R}_k^2\right) \end{cases}$$

Specifically, $\boldsymbol{\omega}_{k-1}$ follows a mixture distribution composed of two Gaussian distributions: the first with a proportion $(1-a)$ and covariance $\mathbf{Q}_{k-1}^1$, and the second with a proportion a and covariance $\mathbf{Q}_{k-1}^2$. Similarly, the distribution of $\boldsymbol{\upsilon}_k$ consists of a Gaussian distribution with a proportion $(1-b)$ and covariance $\mathbf{R}_k^1$, and another with a proportion b and covariance $\mathbf{R}_k^2$. Here,$a = 0.52, b = 0.37$, $\mathbf{Q}_{k-1}^2 = 30\mathbf{Q}_{k-1}^1$ and $\mathbf{R}_k^2 = 50\mathbf{R}_k^1$. $\mathbf{Q}_{k-1}^1$ and $\mathbf{R}_k^1$ are referred to as the uncontaminated Gaussian noise covariance, while $\mathbf{Q}_{k-1}^2$ and $\mathbf{R}_k^2$ are known as the contaminated, higher-valued noise covariance.

The Gaussian mixture model parameters obtained using the improved robust EM algorithm based on our previous work [49] are: $\hat{\alpha}$=0.6, $\hat{a}$=0.55, $\hat{b}$=0.35.

$$\begin{cases} \hat{\mathbf{Q}}_{k-1}^1 = \text{Diag}\left(\begin{bmatrix} 0.11 & 0.20 & 0.09 & 0.18 \end{bmatrix}\right) \\ \hat{\mathbf{Q}}_{k-1}^2 = \text{Diag}\left(\begin{bmatrix} 2.98 & 5.20 & 3.19 & 6.18 \end{bmatrix}\right) \end{cases}$$

$$\begin{cases} \hat{\mathbf{R}}_k^1 = \text{Diag}\left(\begin{bmatrix} 1.11 & 2.20 \end{bmatrix}\right) \\ \hat{\mathbf{R}}_k^2 = \text{Diag}\left(\begin{bmatrix} 47.5 & 99.6 \end{bmatrix}\right) \end{cases}$$

The non-Gaussian noise with heavy tails in the process and measurement is shown in Fig. 4. The ARMSEs of $\mathbf{x}_1$ and $\mathbf{x}_3$ under six filtering algorithms under zero-mean heavy-tailed non-Gaussian noise are shown in Fig. 5. The Table II compares the estimation performance of six estimation algorithms under this noise.

Table II and Fig. 5 show that the proposed algorithm demonstrates the best filtering performance under this zero-mean heavy-tailed non-Gaussian noise. Multiple experiments have shown that the performance of the KF degrades significantly under non-Gaussian noise and is greatly influenced by the initial

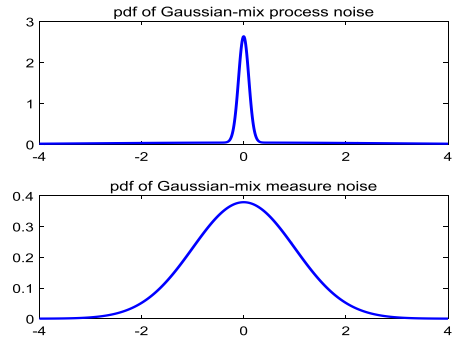

Fig. 4. Heavy-tailed non-Gaussian noise in process and measurement .

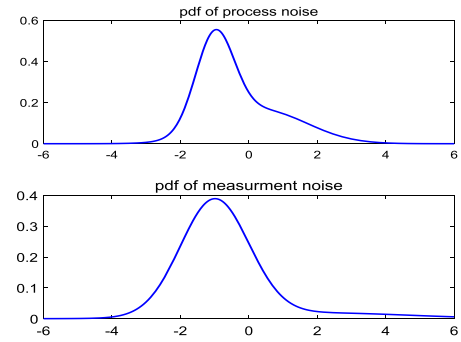

Fig. 6. Bimodal non-Gaussian noise in process noise and measurement noise.

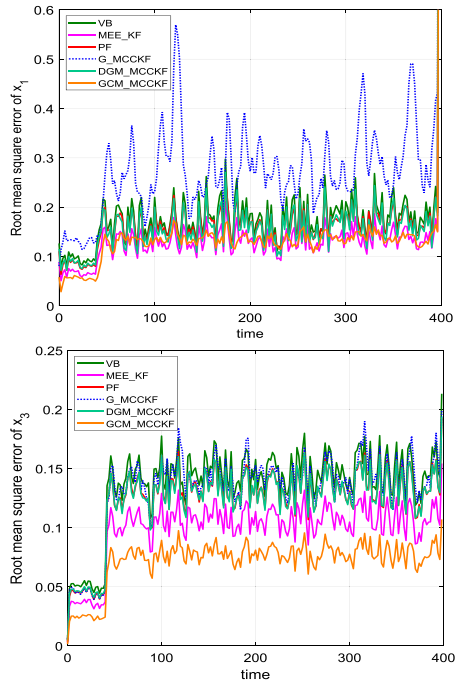

Fig. 5. RMSEs of state variables $\mathbf{x}_1$ and $\mathbf{x}_3$ under heavy-tailed non-Gaussian noise.

TABLE II
ARMSEs OF DIFFERENT FILTERS WITH ZERO-MEAN HEAVY-TAILED NON-GAUSSIAN NOISE

Algorithms	x_1	x_2	x_3	x_4
VB	0.18351	0.19605	0.1366	0.13804
MEE_KF	0.13727	0.14698	0.099651	0.10076
PF	0.17135	0.18346	0.12454	0.12592
G_MCCKF	0.2829	0.30249	0.13301	0.13522
DGM_MCCKF	0.16683	0.17823	0.12418	0.12549
GCM_MCCKF	0.13334	0.14431	0.072222	0.073255

variance $\mathbf{P}_0$; therefore, the performance of the KF under non-Gaussian noise will not be displayed in plots and tables.

3) Scenario 3: The process noise and measurement noise in Equations (15) and (16) are bimodal non-Gaussian heavy-tailed noises, originating from a non-zero-mean Gaussian mixture model. That is,

$$\begin{cases}\boldsymbol{\omega}_{k-1}\sim(1-a)\mathcal{N}\left(\boldsymbol{\mu}_{\omega_{1,k-1}},\mathbf{Q}_{k-1}^{1}\right)+a\mathcal{N}\left(\boldsymbol{\mu}_{\omega_{2},k-1},30\mathbf{Q}_{k-1}^{1}\right)\\ \boldsymbol{\upsilon}_k\sim(1-b)\mathcal{N}\left(\boldsymbol{\mu}_{\upsilon_{1,k}},\mathbf{R}_{k}^{1}\right)+b\mathcal{N}\left(\boldsymbol{\mu}_{\upsilon_{2,k}},50\mathbf{R}_{k}^{1}\right)\end{cases}$$

where $\alpha=0.58, c=20,\ a=0.38, b=0.04$.

$$\begin{cases}\boldsymbol{\mu}_{\omega_{1,k-1}}=\begin{bmatrix}-1 & -1 & -2 & -2\end{bmatrix}\\ \boldsymbol{\mu}_{\omega_{2,k-1}}=\begin{bmatrix}0.4 & 0.4 & 4 & 4\end{bmatrix}\end{cases}$$

$$\begin{cases}\boldsymbol{\mu}_{\upsilon_{1,k}}=\begin{bmatrix}-1 & -3\end{bmatrix}\\ \boldsymbol{\mu}_{\upsilon_{2,k}}=\begin{bmatrix}2 & 2\end{bmatrix}\end{cases}$$

The Gaussian mixture model parameters obtained using the improved robust EM algorithm based on our previous work [49] are: $\hat{\alpha}=0.6$, $\hat{a}=0.35$, $\hat{b}=0.05$.

$$\begin{cases}\hat{\mathbf{Q}}_{k-1}^{1}=\text{diag}\left(\begin{bmatrix}0.10 & 0.22 & 0.13 & 0.21\end{bmatrix}\right)\\ \hat{\mathbf{Q}}_{k-1}^{2}=\text{diag}\left(\begin{bmatrix}2.578 & 5.50 & 3.07 & 5.98\end{bmatrix}\right)\end{cases}$$

$$\begin{cases}\hat{\mathbf{R}}_{k}^{1}=\text{diag}\left(\begin{bmatrix}0.96 & 2.08\end{bmatrix}\right)\\ \hat{\mathbf{R}}_{k}^{2}=\text{diag}\left(\begin{bmatrix}52.6 & 100.3\end{bmatrix}\right)\end{cases}$$

$$\begin{cases}\hat{\boldsymbol{\mu}}_{\omega_{1,k-1}}=\begin{bmatrix}-1.05 & -0.98 & -1.95 & -2.03\end{bmatrix}\\ \hat{\boldsymbol{\mu}}_{\omega_{2,k-1}}=\begin{bmatrix}0.43 & 0.38 & 3.85 & 4.02\end{bmatrix}\end{cases}$$

$$\begin{cases}\hat{\boldsymbol{\mu}}_{\upsilon_{1,k}}=\begin{bmatrix}-1.05 & -0.98\end{bmatrix}\\ \hat{\boldsymbol{\mu}}_{\upsilon_{2,k}}=\begin{bmatrix}1.98 & 2.08\end{bmatrix}\end{cases}$$

The bimodal non-Gaussian noise in process and measurement is shown in Fig. 6. The RMSEs of $\mathbf{x}_1$ and $\mathbf{x}_3$ under six filtering methods under bimodal non-Gaussian noise are shown in Fig. 7. Table III compares the estimation performance of six estimation algorithms with this noise. For the same reasons as in Scenario 2, the performance of the KF in this scenario will not be displayed in the plots and tables.

Table III and Fig. 7 demonstrate that, in contrast to Scenario 2, the performance of all algorithms has decreased in this scenario, with the proposed algorithm showing superior filtering performance.

B) Comparison of filtering performance between fixed kernel size and adaptive kernel size σ in GCM_MCCKF.

Here, we compare the filtering performance under adaptive kernel size and fixed kernel size in heavy-tailed mixed non-Gaussian scenes. The simulation conditions are consistent with

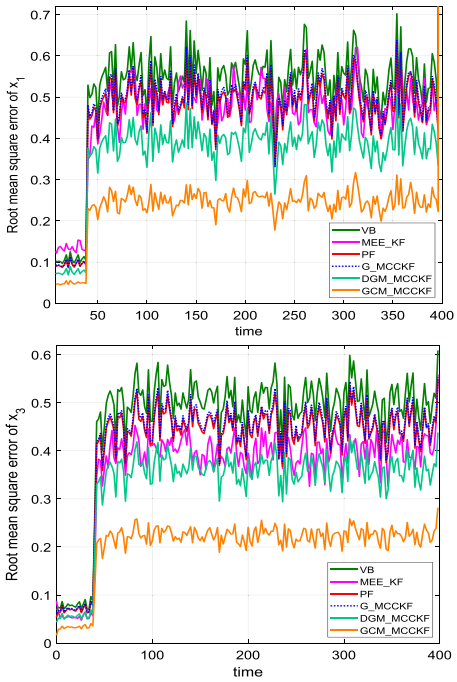

Fig. 7. RMSEs of state variables $\mathbf{x}_1$ and $\mathbf{x}_3$ under bimodal non-Gaussian noise.

TABLE III
ARMSEs OF DIFFERENT FILTERS WITH BIMODAL NON-GAUSSIAN NOISE

Algorithms	ARMSE			
	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
VB	0.74761	0.90824	0.51359	0.56637
MEE_KF	0.69056	0.81678	0.30237	0.34726
PF	0.67105	0.81026	0.45077	0.49652
G_MCCKF	0.67965	0.82567	0.4669	0.51488
DGM_MCCKF	0.53665	0.64788	0.36022	0.39678
GCM_MCCKF	0.38485	0.45913	0.21357	0.23888

Scenario 2. The adaptive kernel size proposed in this paper is shown in Fig. 8. Table IV shows a comparison of simulation performance under different kernel sizes.

It can be seen from the simulation results that the filtering performance was significantly different with different fixed kernel sizes. Filtering performance may not be optimal under the adaptive kernel size rule, but it does not differ significantly from the optimal fixed kernel size that we tested. Therefore, we verified the feasibility of the adaptive kernel size rule. The kernel size value under adaptive rules can provide a reference to set a fixed kernel size. Under this adaptive kernel size rule, we conducted multiple experiments by varying the mixing coefficient α and the heavy-tailed factor c to compare and analyze the data. The conclusions drawn are as follows: 1) With a constant mixing coefficient α, when the heavy-tailed factor is present ($c < 1$), the filter's performance is somewhat unsatisfactory. Setting the heavy-tailed factor greater than 1 improves filtering performance as it increases, reaching a certain threshold beyond which further increments do not significantly

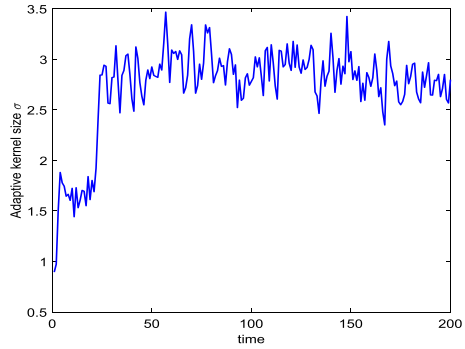

Fig. 8. Adaptive kernel size σ .

TABLE IV
COMPARISON OF SIMULATION PERFORMANCE OF GCM_MCCKF UNDER DIFFERENT KERNEL SIZES

GCM_MCCKF	ARMSE			
	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
$\sigma_{\mathbf{new,k}}$ **in (24)**	0.092752	0.096259	0.06673	0.068161
σ_k **in [52]**	1.16935	1.19213	0.1258	0.13797
$\sigma = 0.5$	64.0737	62.8213	0.507289	0.487305
$\sigma = 1$	20.6611	17.8805	0.198942	0.194955
$\sigma = 5$	0.09005	0.099374	0.068872	0.067573
$\sigma = 10$	0.092705	0.091491	0.067463	0.066995
$\sigma = 100$	0.18093	0.18665	0.092218	0.092099

affect filtering performance. 2) Through multiple experiments by changing the mixing coefficient, we observed that for $c \geq 1$, when $c = 1$, the mixing coefficient has a significant impact on filtering performance. When the heavy-tailed factor is large, the influence of the mixing coefficient on filtering performance is minimal. Both 1) and 2) highlight the importance of the heavy-tailed characteristics of the Cauchy kernel function in the proposed mixed kernel. Furthermore, based on the experimental results of this study, we think that incorporating heavy-tailed factors and mixing coefficients in the design of new adaptive kernel width rules may further improve filtering performance.

C) The influence of the mixing coefficient α in GCM_MCCKF.

Here, we investigate the filtering performance under different mixing coefficients of two kernel functions in GCM_MCCKF under bimodal non-Gaussian noise scenes. The simulation conditions are consistent with Scenario 3 but $c = 2$, $c \neq 20$. Table V shows a comparison of simulation performance under different mixing coefficients.

After repeated experiments, it can be found from the simulation results that the filtering performance of mixture kernel functions is better than that of single Gaussian kernel functions. Similarly to the simulation results in Part B), when the heavy-tail factor is greater than 1 but not very large (e.g., c = 2), different mixing coefficients have a certain impact on the filtering performance, just as Table V shows. However, when the heavy-tail factor is large (e.g. c = 20), different mixing coefficients will achieve approximately equal filtering performance.

TABLE V
COMPARISON OF SIMULATION PERFORMANCE UNDER DIFFERENT KERNEL FUNCTION MIXING COEFFICIENTS α

GCM_MCCKF	ARMSE			
	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
$\alpha = 0$	0.44072	0.60409	0.10753	0.15342
$\alpha = 0.2$	0.28013	0.40151	0.097736	0.13828
$\alpha = 0.4$	0.23868	0.33432	0.093672	0.1314
$\alpha = 0.6$	0.20976	0.30492	0.092536	0.13151
$\alpha = 0.8$	0.21357	0.30038	0.098373	0.13909
$\alpha = 1$	0.23341	0.32986	0.11773	0.16676

D) The influence of the heavy-tailed factor c in GCM_MCCKF.

The objective of this simulation investigation is to examine how heavy-tailed factors c impact filtering performance in bimodal non-Gaussian noise environments. The mixing coefficient of the kernel function here is α = 0.2, and the other simulation conditions are consistent with Scenario 3.

V. CONCLUSION AND FUTURE RESEARCH

In this paper, we proposed the Gaussian-Cauchy mixture maximum correntropy criterion Kalman filter (GCM_MCCKF), designed to optimize the handling of data containing heavy-tailed noise or outliers. Experimentally, through target tracking simulation examples under various noise conditions, this algorithm proves superior in estimation accuracy and filter stability over traditional KF, variational Bayesian filtering (VB), particle filtering (PF), minimum error entropy KF (MEE_KF), single Gaussian kernel MCCKF (G_MCCKF), and double-Gaussian mixture MCCKF (DGM_MCCKF), especially effective against non-Gaussian noise with heavy-tailed or multi-modal characteristics.

Theoretically, our approach integrates features of Gaussian and Cauchy distributions, optimizing a cost function based on information theoretic learning (ITL) to fine-tune filter performance effectively to handle noise with heavy-tailed or multi-modal characteristics. In contrast, Variational Bayesian Filtering optimizes the variational approximation of the posterior probability for non-Gaussian noise, and Particle Filtering approximates the posterior distribution using a large number of samples, which is suitable for nonlinear and non-Gaussian tracking issues.

In practical applications, the Gaussian-Cauchy mixture kernel function requires consideration of not only the kernel size, similar to the Gaussian kernel function, but also the heavy-tail factor c and the mixing coefficient α. This increases computational complexity and may impact real-time processing performance. To mitigate this impact, optimization strategies such as parallel computing, approximate computation, or selective use of the mixture kernel function can be employed to balance performance and real-time requirements. For example, parallel computing can significantly reduce computation time, while approximate computation can reduce computational load while maintaining accuracy. Selective use of the mixture kernel function can be applied at critical moments or under specific conditions to optimize overall performance. We plan to further optimize these methods and perform detailed performance evaluations in our future research.

Future work will also address the current lack of mathematical analysis on the stability of the new algorithm. This paper models non-Gaussian noise using a Gaussian Mixture Model (GMM) and employs an improved Expectation-Maximization (EM) algorithm for iterative parameter estimation of the GMM tailored to the given dataset. Given the real-world context where the covariance of non-Gaussian noise is typically unknown and variable over time, a holistic strategy would be to model or characterize the estimated probability distribution directly, beyond just the mean and variance. This approach might extend to leveraging non-parametric methods for estimating the full probability density function, an avenue we intend to explore in our future work. While the GCM_MCCKF in this study is applied to a linear system, extending its application to nonlinear systems will be considered in future work. Further exploration of adaptive kernel size adjustments will also be a primary focus of future research efforts.

VI. APPENDIX A: DERIVATION OF EQUATIONS (21)-(23)

This appendix provides a detailed derivation of Equations (21)-(23) in the measurement update step.

In the measurement update step, the state estimator $\hat{\mathbf{x}}_k$ is obtained by maximizing the cost function of the Gaussian-Cauchy mixture kernel, which considers both the prediction error term $\|\hat{\mathbf{x}}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1}\|^2_{\mathbf{P}^{-1}_{k|k-1}}$ and the residual term $\|\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k\|^2_{\hat{\mathbf{R}}^{-1}_k}$. The MCC cost function is formed by a convex combination of Gaussian and Cauchy kernels, weighted by a coefficient $\alpha \in [0, 1]$, taking into account the propagation of variance. The cost function is specified as follows:

$$\begin{aligned} M_{GCM,k}(e) &= \alpha G_\sigma + (1-\alpha)C_{c\sigma} \\ &= \alpha(G_{\sigma_r} + G_{\sigma_p}) + (1-\alpha)(C_{c\sigma_r} + C_{c\sigma_p}) \end{aligned} \tag{25}$$

The Gaussian kernel for an error vector e with the kernel size σ and the covariance matrix Σ is defined as follows [37]:

$$G_\sigma(e) = \exp\left(-\frac{\|e\|^2_{\Sigma^{-1}}}{2\sigma^2}\right) = \exp\left(-\frac{e^\top\Sigma^{-1}e}{2\sigma^2}\right)$$

Similarly, the Cauchy kernel function with a kernel size σ, a heavy-tailed factor c, and a covariance matrix Σ can be defined as:

$$C_{c\sigma}(e) = \left(1 + \frac{\|e\|^2_{\Sigma^{-1}}}{c\sigma}\right)^{-1} = \left(1 + \frac{e^\top\Sigma^{-1}e}{c\sigma}\right)^{-1}$$

where $\|e\|^2_{\Sigma^{-1}} = e^\top\Sigma^{-1}e$, and $\|\cdot\|$ represents the 2-norm. For simplicity, the function $G_\sigma(e)$ is abbreviated as G_σ and $C_{c\sigma}(e)$ as $C_{c\sigma}$. We define:

$$\begin{cases} G_{\sigma_r} = G_\sigma\left(\|\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k\|_{\hat{\mathbf{R}}^{-1}_k}\right) \\ G_{\sigma_p} = G_\sigma\left(\|\hat{\mathbf{x}}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1}\|_{\mathbf{P}^{-1}_{k|k-1}}\right) \\ C_{c\sigma_r} = C_{c\sigma}\left(\|\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k\|_{\hat{\mathbf{R}}^{-1}_k}\right) \\ C_{c\sigma_p} = C_{c\sigma}\left(\|\hat{\mathbf{x}}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1}\|_{\mathbf{P}^{-1}_{k|k-1}}\right) \end{cases}$$

Let G_{σ_r} and G_{σ_p} denote the Gaussian kernel functions with kernel size σ for the residuals of the observation equation and the prediction errors of the state equation, respectively. Similarly, the functions $C_{c\sigma_r}$ and $C_{c\sigma_p}$ represent the Cauchy kernel functions, where c adjusts the kernel size σ for the residuals of the observation equation and the prediction errors of the state equation, respectively.

To obtain the optimal state estimation $\hat{\mathbf{x}}_k$, we solve the equation $\frac{\partial M_{GCM,k}(e)}{\partial \hat{\mathbf{x}}_k} = 0$. The detailed expression is as follows:

$$\begin{aligned}\frac{\alpha}{\sigma^2}&\left[G_{\sigma_r}\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}\left(\mathbf{y}_k-\mathbf{H}_k\hat{\mathbf{x}}_k\right)-G_{\sigma_p}\mathbf{P}_{k|k-1}^{-1}\left(\hat{\mathbf{x}}_k-\mathbf{F}_k\hat{\mathbf{x}}_{k-1}\right)\right]\\&+\frac{2(1-\alpha)}{c\sigma}\left[C_{c\sigma_r}^2\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}\left(\mathbf{y}_k-\mathbf{H}_k\hat{\mathbf{x}}_k\right)\right.\\&\left.-C_{c\sigma_p}^2\mathbf{P}_{k|k-1}^{-1}\left(\hat{\mathbf{x}}_k-\mathbf{F}_k\hat{\mathbf{x}}_{k-1}\right)\right]=0\end{aligned}\tag{26}$$

where

$$A_k=\frac{\alpha G_{\sigma_r}}{\sigma^2}+\frac{2(1-\alpha)C_{c\sigma_r}^2}{c\sigma}$$

and

$$B_k=\frac{\alpha G_{\sigma_p}}{\sigma^2}+\frac{2(1-\alpha)C_{c\sigma_p}^2}{c\sigma}$$

are defined. By approximating $\hat{\mathbf{x}}_k\approx\hat{\mathbf{x}}_k^-$ in the Gaussian-Cauchy kernel functions and using (19) in the prediction stage, we have the kernel functions $G_{\sigma_p}\approx 1$, $C_{c\sigma_p}\approx 1$ and

$$B_k=\frac{\alpha}{\sigma^2}+\frac{2(1-\alpha)}{c\sigma}.$$

Substituting the values of A_k and B_k in (26), we obtain the following:

$$A_k\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}(\mathbf{y}_k-\mathbf{H}_k\hat{\mathbf{x}}_k)=B_k\mathbf{P}_{k|k-1}^{-1}(\hat{\mathbf{x}}_k-\hat{\mathbf{x}}_k^-)\tag{27}$$

By adding and subtracting the term $A_k\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}\mathbf{H}_k\hat{\mathbf{x}}_k^-$ to the right side of Equation (27), we derive the state estimate update formula as shown in Equation (28):

$$\hat{\mathbf{x}}_k=\hat{\mathbf{x}}_k^-+\mathbf{K}_{GCM,k}(\mathbf{y}_k-\mathbf{H}_k\hat{\mathbf{x}}_k^-)\tag{28}$$

where $\mathbf{K}_{GCM,k}$ denotes the Kalman filter gain, defined by Equation (29):

$$\begin{aligned}&\mathbf{K}_{GCM,k}\\&=(B_k\mathbf{P}_{k|k-1}^{-1}+A_k\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}\mathbf{H}_k)^{-1}A_k\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}\\&=(\mathbf{P}_{k|k-1}^{-1}+\Psi_{GCM,k}\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}\mathbf{H}_k)^{-1}\Psi_{GCM,k}\mathbf{H}_k^\top\hat{\mathbf{R}}_k^{-1}\end{aligned}\tag{29}$$

here, $\Psi_{GCM,k}$ is the ratio $\frac{A_k}{B_k}$. For simplification, according to the matrix inversion lemma (referenced in [55]), the gain $\mathbf{K}_{GCM,k}$ is commonly expressed in the form shown in Equation (30):

$$\begin{aligned}\mathbf{K}_{GCM,k}=\mathbf{P}_{k|k-1}\Psi_{GCM,k}\mathbf{H}_k^\top(&\mathbf{R}_k\\&+\mathbf{H}_k\mathbf{P}_{k|k-1}\Psi_{GCM,k}\mathbf{H}_k^\top)^{-1}\end{aligned}\tag{30}$$

Based on Equations (28) and (16), we derive the estimation error:

$$\begin{aligned}\tilde{\mathbf{x}}_k&=\mathbf{x}_k-\hat{\mathbf{x}}_k\\&=\mathbf{x}_k-\left(\hat{\mathbf{x}}_k^-+\mathbf{K}_{GCM,k}\left(\mathbf{y}_k-\mathbf{H}_k\hat{\mathbf{x}}_k^-\right)\right)\\&=\mathbf{x}_k-\left(\hat{\mathbf{x}}_k^-+\mathbf{K}_{GCM,k}\left(H_kx_k+\boldsymbol{\upsilon}_k-\mathbf{H}_k\hat{\mathbf{x}}_k^-\right)\right)\\&=(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)\tilde{\mathbf{x}}_k^--\mathbf{K}_{GCM,k}\boldsymbol{\upsilon}_k\end{aligned}\tag{31}$$

where $\mathbf{I}_n$ is the identity matrix of dimension n.

$$\begin{aligned}\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^\top&=[(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)\tilde{\mathbf{x}}_k^--\mathbf{K}_{GCM,k}\boldsymbol{\upsilon}_k]\\&\quad\times[(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)\tilde{\mathbf{x}}_k^--\mathbf{K}_{GCM,k}\boldsymbol{\upsilon}_k]^\top\\&=(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)\tilde{\mathbf{x}}_k^-(\tilde{\mathbf{x}}_k^-)^\top(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)^\top\\&\quad-\mathbf{K}_{GCM,k}\boldsymbol{\upsilon}_k(\tilde{\mathbf{x}}_k^-)^\top(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)^\top\\&\quad-(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)\tilde{\mathbf{x}}_k^-\boldsymbol{\upsilon}_k^\top\mathbf{K}_{GCM,k}^\top\\&\quad+\mathbf{K}_{GCM,k}\boldsymbol{\upsilon}_k\boldsymbol{\upsilon}_k^\top\mathbf{K}_{GCM,k}^\top\end{aligned}\tag{32}$$

Since $\tilde{\mathbf{x}}_k^-$ is a linear function of $\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_{k-1}$, and independent of the current $\mathbf{y}_k$, we have $E(\boldsymbol{\upsilon}_k(\tilde{\mathbf{x}}_k^-)^\top)=0$ and $E(\tilde{\mathbf{x}}_k^-\boldsymbol{\upsilon}_k^\top)=0$. Based on Equation (32), we obtain the estimation error covariance $\mathbf{P}_k$ as follows:

$$\begin{aligned}\mathbf{P}_k&=E(\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^\top)\\&\approx(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I}_n-\mathbf{K}_{GCM,k}\mathbf{H}_k)^\top\\&\quad+\mathbf{K}_{GCM,k}\hat{\mathbf{R}}_k\mathbf{K}_{GCM,k}^\top\end{aligned}\tag{33}$$

The above equation describes the dynamics of estimation error considering measurement noise and process noise. Equations (28), (30) and (33) correspond to Equations (21), (22) and (23) in the measurement update stage of the GCM_MCCKF algorithm derivation, respectively.

REFERENCES

[1] F. Auger, M. Hilairet, J. M. Guerrero, E. Monmasson, T. Orlowska-Kowalska, and S. Katsura, "Industrial applications of the Kalman filter: A review," *IEEE Trans. Ind. Electron.*, vol. 60, no. 12, pp. 5458–5471, Dec. 2013.

[2] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, and J. A. Chambers, "Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 49, no. 10, pp. 2082–2096, Oct. 2019.

[3] J. Perul and V. Renaudin, "BIKES: Bicycle itinerancy Kalman filter with embedded sensors for challenging urban environment," *IEEE Sensors J.*, vol. PP, no. 99, pp. 1–1, Mar. 2022.

[4] Y. Guo, X. Li, and Q. Meng, "A runs test-based Kalman filter with both adaptability and robustness with application to INS/GNSS integration," *IEEE Sensors J.*, vol. 22, no. 23, pp. 22919–22930, Dec. 2022.

[5] E. Ezemobi, M. Silvagni, A. Mozaffari, A. Tonoli, and A. Khajepour, "State of health estimation of lithium-ion batteries in electric vehicles under dynamic load conditions," *Energies*, vol. 15, no. 3, 2022, Art. no. 1234. [Online]. Available: https://www.mdpi.com/1996-1073/15/3/1234

[6] J. Zhao, G. Zhang, Z. Y. Dong, and M. La Scala, "Robust forecasting aided power system state estimation considering state correlations," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2658–2666, Jul. 2018.

[7] J. C. Príncipe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. Springer, 2010.

[8] J. C. Príncipe, D. Xu, J. Fisher, and S. Haykin, "Information theoretic learning: Unsupervised adaptive filtering," in *Unsupervised Adaptive Filtering*, vol. 1, S. Haykin, Ed. Wiley, 2000.

[9] D. Erdogmus and J. C. Príncipe, "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1780–1786, Jul. 2002.

[10] B. Chen, L. Dang, Y. Gu, N. Zheng, and J. C. Príncipe, "Minimum error entropy Kalman filter," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 51, no. 9, pp. 5819–5829, Sep. 2021.

[11] H. Zhao and B. Chen, *Efficient Nonlinear Adaptive Filters: Design, Analysis and Applications*. Springer, 2023, pp. 209–254.

[12] G. T. Cinar and J. C. Príncipe, "Adaptive background estimation using an information theoretic cost for hidden state estimation," in *Proc. Int. Joint Conf. Neural Netw.*, 2011, pp. 489–494.

[13] G. T. Cinar and J. C. Príncipe, "Hidden state estimation using the correntropy filter with fixed point update and adaptive kernel size," in *Proc. 2012 Int. Joint Conf. Neural Netw. (IJCNN)*, 2012, pp. 1–6.

[14] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, and D. Simon, "Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise," in *Proc. Annu. Conf. Inf. Sci. Syst. (CISS)*, 2016, pp. 500–505.

[15] B. Chen, X. Liu, H. Zhao, and J. C. Príncipe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000510981630396X

[16] G. Wang, Y. Zhang, and X. Wang, "Maximum correntropy Rauch–Tung–Striebel smoother for nonlinear and non-Gaussian systems," *IEEE Trans. Autom. Control*, vol. 66, no. 3, pp. 1270–1277, Mar. 2021.

[17] X. Liu, H. Qu, J. Zhao, and B. Chen, "Extended Kalman filter under maximum correntropy criterion," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2016, pp. 1733–1737.

[18] 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.

[19] G. Wang, N. Li, and Y. Zhang, "Maximum correntropy unscented Kalman and information filters for non-Gaussian measurement noise," *J. Franklin Inst.*, vol. 354, no. 18, pp. 8659–8677, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016003217305446

[20] A. Gw, A. Yz, and B. Xw, "Iterated maximum correntropy unscented Kalman filters for non-Gaussian systems," *Signal Process.*, vol. 163, pp. 87–94, 2019.

[21] Z. Deng, L. Shi, L. Yin, Y. Xia, and B. Huo, "UKF based on maximum correntropy criterion in the presence of both intermittent observations and non-Gaussian noise," *IEEE Sensors J.*, vol. 20, pp. 7766–7773, 2020.

[22] J. Shao, W. Chen, Y. Zhang, F. Yu, and J. Chang, "Adaptive multikernel size-based maximum correntropy cubature Kalman filter for the robust state estimation," *IEEE Sensors J.*, vol. 22, no. 20, pp. 19835–19844, Oct. 2022.

[23] 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, 2018, Art. no. S001905781830168X.

[24] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[25] A. Singh and J. Príncipe, "Information theoretic learning with adaptive kernels," *Signal Process.*, vol. 91, no. 2, pp. 203–213, 2011.

[26] C. Hu, G. Wang, K. Ho, and J. Liang, "Robust ellipse fitting with Laplacian kernel based maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 30, pp. 3127–3141, 2021.

[27] H. Dong, L. Yang, and X. Wang, "Robust semi-supervised support vector machines with Laplace kernel-induced correntropy loss functions," *Appl. Intell.*, vol. 51, pp. 819–833, 2021.

[28] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.

[29] Y. Zhu, H. Zhao, X. Zeng, and B. Chen, "Robust generalized maximum correntropy criterion algorithms for active noise control," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1282–1292, 2020.

[30] H. Huang and H. Zhang, "Student's t-kernel-based maximum correntropy Kalman filter," *Sensors*, vol. 22, no. 4, 2022, Art. no. 1683.

[31] J. Wang, D. Lyu, Z. He, H. Zhou, and D. Wang, "Cauchy kernel-based maximum correntropy Kalman filter," *Int. J. Syst. Sci.*, vol. 51, no. 16, pp. 3523–3538, 2020.

[32] S. Zhao, X. Wang, and Y. Liu, "Cauchy kernel correntropy-based robust multi-innovation identification method for the nonlinear exponential autoregressive model in non-Gaussian environment," *Int. J. Robust Nonlinear Control*, 2024.

[33] Y. Wang, Z. Yang, Y. Wang, Z. Li, V. Dinavahi, and J. Liang, "Resilient dynamic state estimation for power system using Cauchy-kernel-based maximum correntropy cubature Kalman filter," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[34] B. Chen, X. Wang, N. Lu, S. Wang, J. Cao, and J. Qin, "Mixture correntropy for robust learning," *Pattern Recognit.: J. Pattern Recognit. Soc.*, 2018.

[35] B. Chen, Y. Xie, X. Wang, Z. Yuan, P. Ren, and J. Qin, "Multikernel correntropy for robust learning," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13500–13511, 2021.

[36] L. Xing, H. Zhao, Z. Lin, and B. Chen, "Mixture correntropy based robust multi-view k-means clustering," *Knowl. Based Syst.*, vol. 262, 2023, Art. no. 110231.

[37] C. Lu, W. Feng, Y. Zhang, and Z. Li, "Maximum mixture correntropy based outlier-robust nonlinear filter and smoother," *Signal Process.*, vol. 188, 2021, Art. no. 108215.

[38] Y. Wang, L. Yang, and Q. Ren, "A robust classification framework with mixture correntropy," *Inf. Sci.*, vol. 491, pp. 306–318, 2019.

[39] H. Wang, W. Zhang, J. Zuo, and H. Wang, "Outlier-robust Kalman filters with mixture correntropy," *J. Franklin Inst.*, pp. 5058–5072, 2019.

[40] H. Wang, H. Li, W. Zhang, J. Zuo, and H. Wang, "Maximum correntropy derivative-free robust Kalman filter and smoother," *IEEE Access*, vol. 6, pp. 70794–70807, 2018.

[41] F. D. Mandanas and C. L. Kotropoulos, "Robust multidimensional scaling using a maximum correntropy criterion," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 919–932, Feb. 2016.

[42] F. Huang, J. Zhang, and S. Zhang, "Adaptive filtering under a variable kernel width maximum correntropy criterion," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 64, no. 10, pp. 1247–1251, Oct. 2017.

[43] W. Wang, J. Zhao, H. Qu, B. Chen, and J. C. Príncipe, "An adaptive kernel width update method of correntropy for channel estimation," in *Proc. IEEE Int. Conf. Digit. Signal Process.*, 2015, pp. 916–920.

[44] W. Wang, J. Zhao, Q. Hua, B. Chen, and J. C. Príncipe, "A switch kernel width method of correntropy for channel estimation," in *Proc. Int. Joint Conf. Neural Netw.*, 2015.

[45] S. Zhao, B. Chen, and J. C. Príncipe, "Kernel adaptive filtering with maximum correntropy criterion," in *Proc. 2011 Int. Joint Conf. Neural Netw.*, 2011, pp. 2012–2017.

[46] J. He, C. Sun, B. Zhang, and P. Wang, "Variational Bayesian-based maximum correntropy cubature Kalman filter with both adaptivity and robustness," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1982–1992, Jan. 2021.

[47] I. Hafez, A. Wadi, M. F. Abdel-Hafez, and A. A. Hussein, "Variational Bayesian-based maximum correntropy cubature Kalman filter method for state-of-charge estimation of Li-ion battery cells," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3090–3104, Mar. 2023.

[48] P. Dong, Z. Jing, H. Leung, and K. Shen, "Variational Bayesian adaptive cubature information filter based on Wishart distribution," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 6051–6057, Nov. 2017.

[49] G. Quan-Bo, W. He-Bin, Y. Qin-Min, Z. Xing-Guo, and L. Hua-Ping, "Estimation of robot motion state based on improved Gaussian mixture model," *Acta Automatica Sinica*, vol. 48, no. 8, pp. 1972–1983, 2022.

[50] B. Chen, J. Liang, N. Zheng, and J. C. Príncipe, "Kernel least mean square with adaptive kernel size," *Neurocomputing*, vol. 191, pp. 95–106, 2014.

[51] S. Zhao, B. Chen, and J. C. Príncipe, "An adaptive kernel width update for correntropy," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2012, pp. 1–5.

[52] B. Hou, Z. He, X. Zhou, H. Zhou, and J. Wang, "Maximum correntropy criterion Kalman filter for α-jerk tracking model with non-Gaussian noise," *Entropy*, vol. 19, no. 12, 2017, Art. no. 648.

[53] H. Zhu, H. Leung, and Z. He, "State estimation in unknown non-Gaussian measurement noise using variational Bayesian technique," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 4, pp. 2601–2614, Oct. 2013.

[54] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[55] D. Simon, *Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches*. New York, NY, USA: Wiley-Interscience, 2006.

Quanbo Ge (Member, IEEE) received the bachelor's and master's degrees from the College of Computer and Information Engineering, Henan University, Kaifeng, China, in 2002 and 2005, respectively, and the Ph.D. degree from Shanghai Maritime University, Shanghai, China, in 2008. He was a Professor with the Institute of Systems Science and Control Engineering, School of Automation, Hangzhou Dianzi University, Hangzhou, China. From 2008 to 2010, he was a Lecturer and became an Associate Professor with the School of Automation, Hangzhou Dianzi University. From 2009 to 2013, he was a Postdoctoral Fellow with the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, China. From 2012 to 2013, he was a Visiting Scholar with the Optimization for Signal Processing and Communication Group, Department of Electrical and Computer Engineering, Twin Cities Campus, University of Minnesota, Minneapolis, MN, USA. Currently, he is a Professor with the School of Automation, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include information fusion, autonomous unmanned systems, man-mechanic hybrid systems, and machine vision.

Xuefei Bai (Student Member, IEEE) received the bachelor's degree in electrical engineering and automation from Heilongjiang University of Science and Technology, in 2005, and the master's degree in power electronics and power transmission from Jiangsu University, in 2008. She is currently working toward the Ph.D. degree in control science and engineering from Hangzhou Dianzi University. Her research interests include filtering algorithms and information fusion.

Pingliang Zeng (Senior Member, IEEE) received the bachelor's degree in electrical engineering from Huazhong University of Science and Technology, China, in 1984, and the Ph.D. degree in electrical engineering from Strathclyde University, U.K., in 1990. He joined National Grid Company (NGC), U.K., where he held several important positions, including a Chartered Engineer (CEng), a System Operation Incentive Mechanism and Strategy Department Manager, a Transmission Network Design Department Manager, etc. In 2012, he joined as the Chief Expert in power system analysis and planning with the China Electric Power Research Institute, where he led a research team carrying out several key projects, including National High Technology Development ("863" Plan), SGCC special projects, and key Sino-UK joint research projects funded by the China National Natural Science Foundation on the integration of EV and energy storage in power systems. He joined Hangzhou Dianzi University as a Professor in 2017. He has published 2 books and about 130 academic papers. He is a National Special Expert, a fellow of the IET, and a member of IEC TC122 and TC8. His research interests include power system analysis and planning, renewable energy grid connection, electric vehicle access, power system reliability, large-scale energy storage applications, distributed power generation, load side response, and comprehensive energy systems.