Maximum Correntropy Derivative-Free Robust Kalman Filter and Smoother

HONGWEI WANG[1,2], **(Student Member, IEEE), HONGBIN LI**[2], **(Senior Member, IEEE),**
WEI ZHANG[1], **JUNYI ZUO**[1], **AND HEPING WANG**[1]

[1]School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China
[2]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

Corresponding author: Wei Zhang (weizhangxian@nwpu.edu.cn)

ABSTRACT We consider the problem of robust estimation involving filtering and smoothing for nonlinear state space models which are disturbed by heavy-tailed impulsive noises. To deal with heavy-tailed noises and improve the robustness of the traditional nonlinear Gaussian Kalman filter and smoother, we propose in this work a general framework of robust filtering and smoothing, which adopts a new maximum correntropy criterion to replace the minimum mean square error for state estimation. To facilitate understanding, we present our robust framework in conjunction with the cubature Kalman filter and smoother. A half-quadratic optimization method is utilized to solve the formulated robust estimation problems, which leads to a new maximum correntropy derivative-free robust Kalman filter and smoother. Simulation results show that the proposed methods achieve a substantial performance improvement over the conventional and existing robust ones with slight computational time increase.

INDEX TERMS Robust Kalman filtering, robust Kalman smoothing, maximum correntropy criterion, heavy-tailed noise, half-quadratic minimization.

I. INTRODUCTION

In recent years, estimation problems involving filtering and smoothing based on dynamic state space models (SSMs) have received significant attentions. These problems are frequently encountered in many areas, such as target tracking, fault detection and diagnosis, parameter estimation, navigation, and many others. The celebrated Kalman filter (KF) [1] offers optimal estimation with the minimum mean square error (MMSE) for a linear SSM when both the process and measurement noises are Gaussian. Nevertheless, most applications in practice inherently have nonlinear SSMs, for which an optimal nonlinear filter or smoother is typically intractable. Several sub-optimal solutions were proposed for nonlinear SSMs, including the extended KF (EKF) [2], unscented KF (UKF) [3], cubature Kalman filter (CKF) [4], and others. Some interesting relations among these solutions can be found in [5] and [6]. Meanwhile, for smoothing, the Rauch-Tung-Striebel (RTS) smoother was introduced in [7] for linear Gaussian SSMs. A general framework of Gaussian optimal smoothing for nonlinear SSMs was proposed in [8], following which several sub-optimal nonlinear RTS smoothers were developed, such as the extended RTS smoother, unscented RTS smoother, cubature Kalman smoother (CKS), and others.

Although the aforementioned filtering and smoothing methods in general perform well under the Gaussian assumption, they can potentially break down in the presence of heavy-tailed non-Gaussian noises, which may appear in either the process procedure or measurement procedure. This happens in, for instance, tracking a maneuvering target with outliers in observations [9]. A primary reason behind the degradation is that the traditional KF/RTS and their sub-optimal extensions were derived from the MMSE criterion, and the resulting estimates are sensitive to heavy-tailed noises [10].

Persistent efforts have been devoted to tackling heavy-tailed noises in order to obtain more robust state estimates. Multiple models based techniques and sequential Monte Carlo sampling methods can be used to handle non-Gaussian noises [11], [12]. Unfortunately, their heavy computational burden makes them considerably more difficult to implement in real time. Another common strategy to enhance

the robustness is robust statistics (or influence functions). Masreliez and Martin [13] introduced the Huber cost function [14] to a linear SSM to construct a robust KF by recasting the filtering problem as a linear regression. After that, this type of robust filters were extended to nonlinear SSMs via linearization [15]. More recently, to reduce the approximation error arising from linearization, nonlinear regression based Huber's robust filters were proposed in [16] and [17]. Other approaches for robust estimation, e.g., the H_∞ filter [18] and heavy-tailed distributions based filters and smoothers [19]–[21], were also studied.

In addition, several recent studies explored using optimization to design robust smoothers since smoothing is in essence an optimization problem. Specifically, an ℓ_1 robust Kalman smoother was presented in [22] based on an interior point method, where the Laplace distribution was utilized to model the measurement noise. Quadratic programming was employed to design a student's t Kalman smoother in [23], where both the process and measurement noise were modeled by student's t distribution. Quadratic support functions were introduced in [24] to present generalized Kalman smoothing methods. Other optimization methods for robust Kalman smoothing were discussed in [25]. Although optimization based methods provide reasonable smoothing results, they still have some limitations. A major one is that the methods ignore covariance propagation, and hence some important statistic information about the state may get lost. Besides, these methods were primarily designed for smoothing, and it is difficult to extend them for filtering, which is more frequently encountered in real-time applications.

Recently a concept called correntropy in information theoretic learning and an associated maximum correntropy criterion (MCC) [26] was utilized as a new cost function for robust filtering [27]. As shown in [26], MCC has a close relationship with M-estimators. In fact, correntropy is a robust recasting of the Welsch cost function [28]. Compared with Huber's cost function, the Welsch function penalizes a large fitting errors [26]. Furthermore, the Welsch M-estimator outperforms Huber's M-estimator in dealing with different types of outliers. Meanwhile the kernel size in the Welsch is less sensitive to select than the threshold in Huber's M-estimator, which can have a significant impact on the estimation performance [29].

In the present paper, we focus on developing a general robust framework, including both robust filtering and robust smoothing, to estimate the states of nonlinear SSMs where heavy-tailed noises may be present in both the process and measurement procedures. The proposed framework is based on the MCC and a half-quadratic (HQ) optimization method. Numerical results show that the proposed algorithms can provide reliable state estimation, and are robust to almost all levels of contamination. In comparison with existing robust algorithms, our methods yield more accurate results with a moderate increase of computation time.

A. RELATED WORK

Correntropy and its associated MCC criterion was originally proposed in [26], where the relationship between MCC and M-estimation was established. The advantages of the MCC in dealing with non-Gaussian noises with heavy tails were demonstrated in [26] via numerical simulations. Thereafter, the MCC criterion was employed to improve the robustness in several areas, such as feature selection [30], robust face recognition [31], principal component analysis [32], and so on. During last few years, several researchers started to employ the MCC to deal with heavy-tailed noises in Kalman filtering. It was first utilized to improve the robustness of the KF in [33] and [34], where gradient descent was used for state estimation. Those methods, however, ignored covariance propagation, which is an important part in the KF.

In [35], a robust KF called MCC-KF was proposed, which integrated both the MCC and weighted least squares into the traditional KF framework. A square-root format of the MCC-KF was proposed in [36] to improve the numerical stability. MCC-KF and its square-root version only modified the weights of the measurements and thus had the same prediction step as the traditional KF which was derived under the Gaussian assumption. Consequently, the MCC-KF and its square-root version were sensitive to non-Gaussian process noise. To handle this issue, another robust KF called MCKF [37] was developed using an approach similar to that of the Huber based robust KF, i.e., recasting the filtering problem as a linear regression. In MCKF, a fixed-point algorithm was employed to iteratively provide robust estimation, and both the prediction step and measurement update were adapted. Henceforth, several variants of the MCKF were proposed for nonlinear systems by linearizing the nonlinear mappings, e.g., the Taylor series [38], unscented transformation [27], and Gauss-Hermite quadrature rule [39]. Most recently, the MCC was employed to solve the robust estimation problem in the continuous-discrete system [40]. In addition, the MCC was also utilized to suppress outliers in measurements assuming Gaussian noise in the process [41], [42].

Although above studies have clearly demonstrated the potential of the MCC for robust Kalman filtering, there are some deficiencies with the state-of-the-art research in this area. Specifically, a general framework of using MCC for robust Kalman filtering and smoothing is missing. Most existing works have addressed the problem under a specific scenario, but the resulting methods may degrade significantly under a different scenario. For example, the method in [27] was developed under the assumption that the process and measurement are both contaminated, its performance may experience considerable degradation under a different scenario, e.g., when only the process or the measurement is contaminated. In addition, most recently developed robust Kalman filtering schemes using the MCC were based on linear regression [27], [37]–[39], where a linearization procedure is inevitable, resulting in a loss of accuracy.

Another limitation is that no existing work has explored MCC for robust Kalman smoothing.

B. MAIN CONTRIBUTIONS

The main contributions of this paper are summarized as follows.

- We propose a general framework that integrates MCC for robust Kalman filtering. We take an optimization perspective based approach, rather than linear regression as employed in most existing MCC based robust Kalman filtering solutions. This approach bypasses the linearization procedure, which not only makes our methods derivative-free but also reduces approximation errors caused by linearization. Our proposed solutions are derivative-free and, hence, enjoy the benefit of requiring no analytical Jacobians of the SSMs.
- A distinctive feature of the proposed framework involves replacing the quadratic cost in the traditional KF by a sum of weighted instantaneous correntropy terms to improve robustness. The weight of each correntropy is automatically tuned to cope with different scenarios involving Gaussian or heavy-tailed non-Gaussian noises.
- The optimization problem underlying the proposed framework is solved by a half-quadratic method. This enables our robust Kalman filtering algorithm to be cast using the conventional KF iterative procedures, thus inheriting benefits associated with the latter.
- We also extend the above robust filtering framework for robust smoothing, resulting in a MCC based robust Kalman smoother. To the best of our knowledge, our approach is the first robust nonlinear Kalman smoother based on the MCC.

The paper is organized as follows. We first formulate the problem in Section II, and derive the robust nonlinear Kalman filter based on MCC in Section III. In Section IV, the robust nonlinear Kalman smoother is presented. We show numerical examples to illustrate the performance of the proposed robust filtering and smoothing algorithms in Section V. Finally, we draw conclusions and summarize future related research in Section VI.

Notations: In this paper, normal font letters are scalars, boldface lowercase letters denote column vectors, and boldface uppercase letters mean matrices. The estimate of state x_t given the measurements from $t = 1$ to $t = m$ is denoted by $\hat{x}_{t|m}$ where $m = t$ means filtering, and $m = T$ means smoothing, with T denoting the total number of observations. I_n denotes the n-dimensional identity matrix. $\mathcal{N}(\cdot, \cdot)$ denotes a Gaussian distribution.

II. PROBLEM FORMULATION

In this paper, we address a sequential estimation problem in nonlinear systems involving non-Gaussian noises. More specifically, both the process noise and measurement noise are primarily Gaussian random processes; but may occasionally be contaminated by outliers with an unknown distribution. To model this, we consider the following nonlinear discrete-time state-space model (SSM)

$$x_t = f(x_{t-1}) + (1 - \rho_1)v_{t-1} + \rho_1 \varepsilon_1 \qquad (1)$$
$$y_t = h(x_t) + (1 - \rho_2)w_t + \rho_2 \varepsilon_2 \qquad (2)$$

where $t = \{1, \cdots, T\}$ is the time index; $x_t \in \mathbb{R}^n$ is the state of interest to be estimated; $y_t \in \mathbb{R}^m$ is the observation; $f(\cdot)$ and $h(\cdot)$ are some known nonlinear mappings, representing the state transition and measurement procedure, respectively; v_{t-1} and w_t are zero-mean Gaussian distribution vectors with Q_{t-1} and R_t as covariance, representing the nominal process noise and observation noise, respectively; ε_1 and ε_2 are some contaminating noises; ρ_1 and ρ_2 are the corresponding contamination ratios. The initial state x_0 is assumed to follow a known Gaussian prior distribution $p(x_0) = \mathcal{N}(\hat{x}_{0|0}, P_{0|0})$. Furthermore, $x_0, v_{t-1}, w_t, \varepsilon_1$ and ε_2 are assumed to be mutually independent.

Practically, the distributions of the contaminating noises and their corresponding contamination ratios are often unknown. If one chooses to ignore the contaminating noises and run a Gaussian approximation filter and smoother, the performance will likely to be unsatisfying or even diverge [14]. Although ineffective in dealing with the contaminating noise, the Gaussian approximation filter and smoother are popular in those nonlinear sequential estimation when both process and measurement noises are Gaussian due to their computational efficiency and relative high accuracy. This motivates us to explore extended versions of the Gaussian approximation filter and smoother to deal with the contaminating noise in nonlinear systems.

The aim of this work is to develop a unified framework for robust nonlinear filter and smoother design with the information $p(x_0), p(v_t), p(w_t)$ and the observations (namely, $y_{1:t}$ for filtering and $y_{1:T}$ for smoothing, respectively). To address non-Gaussian contaminating noise, the proposed framework employs a new maximum correntropy criterion (MCC) [26] which integrated with the Gaussian approximation filter and smoother, resulting in a class of MCC based filter and smoother solutions. The developed robust filtering and smoothing techniques are expected to have a similar performance as the Gaussian approximation counterparts in the absence of outliers, meanwhile exhibiting less sensitivity in scenarios involving contaminating noise.

In the following, our robust framework is presented in conjunction with the CKF [4] and CKS [43], which are popular for nonlinear system estimation. A brief summary of the CKF and CKS is included in Appendix A to facilitate the presentation of our framework. It is straightforward to extend the framework for use with other nonlinear Gaussian approximation filters and smoothers.

III. ROBUST NONLINEAR KALMAN FILTER WITH MCC

A. MCC-BASED FILTERING

In the Bayesian filtering paradigm, the core problem is to calculate the posterior density through Bayes' rule based on

the measurements up to t,

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) \quad (3)$$

where $p(\boldsymbol{y}_t|\boldsymbol{x}_t)$ is the likelihood function which is defined by the measurement process and $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:t-1})$ is the predictive density which is controlled by Chapman-Kolmogorov equation

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})d\boldsymbol{x}_{t-1} \quad (4)$$

where $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ is the posterior density at time $t-1$ and $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ is the transition density which is defined by the state transition process. The maximum a posteriori estimation (MAP) of $\boldsymbol{x}_t$ can be obtained through minimizing the negative log posterior density, i.e.,

$$\begin{aligned} \hat{\boldsymbol{x}}_{t|t} &= \arg\min_{\boldsymbol{x}_t} -\log p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) \\ &= \arg\min_{\boldsymbol{x}_t} \Big(-\log p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) - \log p(\boldsymbol{y}_t|\boldsymbol{x}_t) \Big) \end{aligned} \quad (5)$$

Under the Gaussian approximation, i.e., both the process noise and measurement noise are assumed Gaussian and the predictive density is a Gaussian distribution, dropping terms that do not depend on $\boldsymbol{x}_t$, we can rewrite (5) as [17]

$$\hat{\boldsymbol{x}}_{t|t} = \arg\min_{\boldsymbol{x}_t} \left(\frac{1}{2}\|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}\|^2_{\boldsymbol{P}^{-1}_{t|t-1}} + \frac{1}{2}\|\boldsymbol{y}_t - h(\boldsymbol{x}_t)\|^2_{\boldsymbol{R}^{-1}_t} \right) \quad (6)$$

where $\|\boldsymbol{z}\|^2_{\boldsymbol{A}} = \boldsymbol{z}^T\boldsymbol{A}\boldsymbol{z}$ and $\hat{\boldsymbol{x}}_{t|t-1}$ is the predicted state with the prediction error covariance $\boldsymbol{P}_{t|t-1}$. We define the normalized error as

$$\boldsymbol{\alpha}_t = \boldsymbol{P}^{-1/2}_{t|t-1}\left(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1}\right) \quad (7)$$

$$\boldsymbol{\beta}_t = \boldsymbol{R}^{-1/2}_t\left(\boldsymbol{y}_t - h(\boldsymbol{x}_t)\right) \quad (8)$$

Then (6) can be further interpreted as

$$\begin{aligned} \hat{\boldsymbol{x}}_{t|t} &= \arg\min_{\boldsymbol{x}_t} \left(\frac{1}{2}\boldsymbol{\alpha}_t^T\boldsymbol{\alpha}_t + \frac{1}{2}\boldsymbol{\beta}_t^T\boldsymbol{\beta}_t \right) \\ &= \arg\min_{\boldsymbol{x}_t} \left(\frac{1}{2}\sum_{i=1}^n \alpha_{t,i}^2 + \frac{1}{2}\sum_{j=1}^m \beta_{t,j}^2 \right) \end{aligned} \quad (9)$$

where $\alpha_{t,i}$ is the i-th component of $\boldsymbol{\alpha}_t$ and $\beta_{t,j}$ is the j-th component of $\boldsymbol{\beta}_t$.

Equation (9) shows that the Gaussian approximation filtering approach is actually a minimum mean square error (MMSE) estimator. It is clear that the quadratic function (i.e., ℓ_2 norm) in MMSE has an apparent effect of emphasizing the contribution of large errors, which makes the MMSE estimator sensitive to impulsive noise. To improve the robustness, the ℓ_2-norm objective function should be replaced by another cost function which is insensitive to the impulsive noise.

We propose to integrate the concept of correntropy [26], which has been shown effective in dealing with outliers [44], for robust Kalman filtering and smoothing, by utilizing the maximum correntropy criterion (MCC) to replace the MMSE criterion. Specifically, we apply the instantaneous correntropy to each component of the normalized error, resulting in the following cost function for robust nonlinear filtering:

$$\hat{\boldsymbol{x}}_{t|t} = \arg\max_{\boldsymbol{x}_t} \left(\sum_{i=1}^n a_{t,i}\kappa_{\sigma_{t,i}}(\alpha_{t,i}) + \sum_{j=1}^m b_{t,j}\kappa_{\eta_{t,j}}(\beta_{t,j}) \right) \quad (10)$$

where $\kappa_\sigma(\cdot)$ is the kernel function which satisfies Mercer's Theorem [45]; $a_{t,i}$ and $b_{t,j}$ are weighting coefficients to be further determined. The kernel function plays a central role in MCC and different kernel functions provide different estimation results.

Without loss of generality, we apply the the Gaussian kernel function with bandwidth σ,

$$\kappa_\sigma(e) = \exp\left(-\frac{e^2}{2\sigma^2}\right).$$

The Gaussian kernel based MCC is closely related to the Welsch M-estimator, where outliers are given small weights in optimization [29]. For completeness, in Appendix B, we provide a short review of the correntropy, MCC, and related properties.

There are two aspects that should be noted in (10). First, a weighting coefficient (namely $a_{t,i}$ and $b_{t,j}$) is introduced to each correntropy. These coefficients should be properly selected to maintain consistency to the Gaussian approximation filter when the SSM does not involve impulse noises (see Section III-B for details on selection for these parameters). Second, having distinct bandwidth for each correntropy offers additional flexibility of adaptiveness to different applications. One possible strategy is to use the same kernel bandwidth for each component of the predictive error and a different kernel bandwidth for the measurement error terms in (10). This makes the proposed algorithms suitable for handling different process noise and measurement noise. For example, in a case where outliers only occur in the measurement procedure, we can select a large σ_t so that the resulting MCC behaves like a quadratic criterion, and a smaller η_t to defend the outliers.

B. A HALF-QUADRATIC OPTIMIZATION BASED SOLUTION
A number of optimization techniques such as the steepest descent [26] and the fixed point iteration approach [46] can in principle be applied to solve MCC based problems. However, our correntropy criterion (10) involves different kernel bandwidths. The problem can be more efficiently solved by a half-quadratic approach [47]. Specifically, our proposed solution to (10) applies the half-quadratic approach in an iterative manner, through which auxiliary variables are introduced to transform the original problem into a quadratic problem. The resulting quadratic problem can be cast in the conventional Gaussian approximation filtering framework. This is a major advantage of our proposed solution. The half-quadratic approach is based on the concept of the convex conjugate function and, in particular, the following result.

Proposition 1: There exists a convex function ψ: $[-1, 0) \rightarrow [-1, 0)$, such that

$$\kappa_\sigma(x) = \sup_{-1 \leq p < 0} \left(\frac{p}{2\sigma^2} x^2 - \psi(p) \right) \tag{11}$$

and for a fixed x, the supremum is reached at $p = -\kappa_\sigma(x)$.

Proof: see Appendix C.

Using (11), the robust filtering problem (10) can be reinterpreted in an augmented format as

$$\hat{x}_{t|t} = \arg\max_{p_t, q_t, x_t} \left\{ \sum_{i=1}^{n} a_{t,i} \left(\frac{p_{t,i}}{2\sigma_{t,i}^2} \alpha_{t,i}^2 - \psi(p_{t,i}) \right) \right.$$
$$\left. + \sum_{j=1}^{m} b_{t,j} \left(\frac{q_{t,j}}{2\eta_{t,j}^2} \beta_{t,j}^2 - \psi(q_{t,j}) \right) \right\} \tag{12}$$

where $p_t = \{p_{t,i}\}_{i=1}^{n}$ and $q_t = \{q_{t,j}\}_{j=1}^{m}$ are the collections of the auxiliary variables. As shown in Section III-C, maximizing (10) with respect to x_t is equivalent to maximizing the augmented cost function (12) in the enlarged parameter domain, i.e., $\{p_t, q_t, x_t\}$. The augmented optimization problem can be solved in an alternating maximization manner. Specifically, we first maximize (12) with respect to p_t and q_t while keeping x_t fixed. Then x_t is updated with the newest p_t and q_t. This process is repeated until a convergence criterion is met.

For the first step of the k-th iteration after x_t^{k-1} is given and $\{\alpha_t^{k-1}, \beta_t^{k-1}\}$ are computed via $\{(7), (8)\}$, by Proposition 1, p_t^k and q_t^k are updated as follows,

$$p_{t,i}^k = -\exp\left(-\frac{(\alpha_{t,i}^{k-1})^2}{2\sigma_{t,i}^2} \right) \tag{13}$$

$$q_{t,j}^k = -\exp\left(-\frac{(\beta_{t,j}^{k-1})^2}{2\eta_{t,j}^2} \right) \tag{14}$$

At the beginning of the iterative procedure, i.e., $k = 1$, we set $\alpha_t^0 = \beta_t^0 = 0$ as the initial condition.

For the second step of the k-th iteration, we fix $\{p_t^k, q_t^k\}$ and drop the items that do not depend on x_t, and then (12) can be simplified as

$$\hat{x}_{t|t}^k = \arg\max_{x_t} \left(\sum_{i=1}^{n} \frac{a_{t,i} p_{t,i}^k}{2\sigma_{t,i}^2} \alpha_{t,i}^2 + \sum_{j=1}^{m} \frac{b_{t,j} q_{t,j}^k}{2\eta_{t,j}^2} \beta_{t,j}^2 \right) \tag{15}$$

Before optimizing the above cost function, it is necessary to discuss the choices of the constants $a_{t,i}$ and $b_{t,j}$. In order to maintain the consistency of the Gaussian approximation filter, $a_{t,i}$ and $b_{t,j}$ should be properly selected so that (15) is equivalent to (9) when both process and measurement noises in the SSM are Gaussian. Specifically, when both noises follow the Gaussian distribution, the following constraints must be in place

$$\frac{a_{t,i} p_{t,i}^k}{\sigma_{t,i}^2} = -1 \tag{16}$$

$$\frac{b_{t,j} q_{t,j}^k}{\eta_{t,j}^2} = -1 \tag{17}$$

In such a case, all bandwidths should be chosen as large as possible so that the correntropy behaves like the ℓ_2 norm, and it follows from Proposition 1 that we can approximate $\{p_{t,i}^k, q_{t,j}^k\}$ by constant -1. Then the coefficients are obtained as follows

$$a_{t,i} = \sigma_{t,i}^2, \tag{18}$$

$$b_{t,j} = \eta_{t,j}^2 \tag{19}$$

Substituting (18) and (19) into (15) leads to the updated cost function:

$$\hat{x}_{t|t}^k = \arg\max_{x_t} \left(\sum_{i=1}^{n} \frac{p_{t,i}^k}{2} \alpha_{t,i}^2 + \sum_{j=1}^{m} \frac{q_{t,j}^k}{2} \beta_{t,j}^2 \right) \tag{20}$$

which can equivalently expressed in a matrix format as

$$\hat{x}_{t|t}^k = \arg\min_{x_t} \left(\frac{1}{2} \alpha_t^T \Psi_t \alpha_t + \frac{1}{2} \beta_t^T \Phi_t \beta_t \right) \tag{21}$$

where $\Psi_t \in \mathbb{R}^{n \times n}$ and $\Phi_t \in \mathbb{R}^{m \times m}$ are diagonal matrices given by

$$\Psi_t = \text{diag}(-p_{t,1}^k, \cdots, -p_{t,n}^k) \tag{22}$$

$$\Phi_t = \text{diag}(-q_{t,1}^k, \cdots, -q_{t,m}^k) \tag{23}$$

Furthermore, substituting (7) and (8) into (21), we obtain

$$\hat{x}_{t|t}^k = \arg\min_{x_t} \left(\frac{1}{2} \|x_t - \hat{x}_{t|t-1}\|_{\bar{P}_{t|t-1}^{-1}}^2 \right.$$
$$\left. + \frac{1}{2} \|y_t - h(x_t)\|_{\bar{R}_t^{-1}}^2 \right) \tag{24}$$

where

$$\bar{P}_{t|t-1} = P_{t|t-1}^{1/2} \Psi_t^{-1} P_{t|t-1}^{T/2} \tag{25}$$

$$\bar{R}_t = R_t^{1/2} \Phi_t^{-1} R_t^{T/2} \tag{26}$$

We notice that, expect for a different weighting matrix, (24) is similar to (6), which can be efficiently solved in the Kalman filtering framework. Hence the second step of the k-th iteration can be implemented in the Kalman filtering framework with the modified weighting matrices. Once the $\hat{x}_{t|t}^k$ is obtained, we can update the weighting matrices and then solve the problem to obtain $\hat{x}_{t|t}^{k+1}$. The iterative procedure is repeated until a convergence criterion is met. The resulting robust cubature Kalman filter (RCKF) is summarized in Algorithm 1.

A closer examination of the proposed algorithm reveals that it is similar to the conventional Gaussian approximation Kalman filter, with the same prediction update step and iterative measurement update step. As proven by simulation results, the iteration number is relatively small in general (see Section V for more details). Hence the proposed MCC based robust nonlinear Kalman filter has only a slightly higher computational complexity compared with the corresponding Gaussian approximation filer.

Algorithm 1 Robust Cubature Kalman Filter Based on MCC

Input: $y_{1:T}, \hat{x}_{0|0}, P_{0|0}, Q_{1:T}, R_{1:T}$.
Output: $\hat{x}_{t|t}$ and $P_{t|t}$ for $t = 1 : T$.
for $t = 1, \cdots, T$ **do**

   1. Run the prediction step of CKF to calculate $\hat{x}_{t|t-1}$ and $P_{t|t-1}$;

   2. Initialize $\Psi_t^0 = I_n, \ \Phi_t^0 = I_m, \ k = 0$;

  repeat

     1. Update $\bar{P}_{t|t-1}$ via (25) and $\bar{R}_t$ via (26);

     2. Update the innovation covariance P_{yy} and the filtering gain K_t via (A.6) and (A.8);

     3. Calculate the filtering state $\hat{x}_{t|t}^k$ via (A.4);

     4. Update Φ_t and Ψ_t via (7), (8),(13), (14), (22) and (23);

     5. $k = k + 1$;

  until a convergence criterion is met

  $\hat{x}_{t|t} = \hat{x}_{t|t}^{k-1}; \ P_{t|t} = P_{t|t-1} - K_t P_{yy} K_t^T.$

end for

C. CONVERGENCE

In this section, we provide a convergence analysis for the MCC based robust nonlinear Kalman filter. For brevity, we denote the original objective function (10) and the transformed cost function (12) by $\mathcal{Q}(x_t)$ and $\mathcal{J}(x_t, p_t, q_t)$, respectively. We first show that $\mathcal{Q}(x_t)$ and $\mathcal{J}(x_t, p_t, q_t)$ have the same optimal solution. We assume that x_t^+ and $\{x_t^*, p_t^*, q_t^*\}$ are the optimal solutions to $\mathcal{Q}(x_t)$ and $\mathcal{J}(x_t, p_t, q_t)$, respectively; in addition, $\{p_t^+, q_t^+\}$ maximize $\mathcal{J}(x_t, p_t, q_t)$ when x_t equals x_t^+. It is clear that

$$\mathcal{Q}(x_t^+) \geq \mathcal{Q}(x_t^*) \tag{27}$$

$$\mathcal{J}(x_t^*, p_t^*, q_t^*) \geq \mathcal{J}(x_t^+, p_t^+, q_t^+) \tag{28}$$

According to Proposition 1, the following equalities hold

$$\mathcal{Q}(x_t^+) = \mathcal{J}(x_t^+, p_t^+, q_t^+) \tag{29}$$

$$\mathcal{Q}(x_t^*) = \mathcal{J}(x_t^*, p_t^*, q_t^*) \tag{30}$$

Combing (27)-(30), we obtain that $\mathcal{Q}(x_t^+) = \mathcal{Q}(x_t^*)$, which illustrates that the original optimization problem (10) and the transformed one (12) have the same optimal solution.

Next, we prove that Algorithm 1 converges, i.e.,

$$\lim_{k \to \infty} \left(\mathcal{J}(x_{t|t}^{k+1}, p_t^{k+1}, q_t^{k+1}) - \mathcal{J}(x_{t|t}^k, p_t^k, q_t^k) \right) = 0 \tag{31}$$

According to Algorithm 1, the following equations hold

$$\mathcal{J}(x_{t|t}^k, p_t^{k+1} q_t^{k+1}) \geq \mathcal{J}(x_{t|t}^k, p_t^k, q_t^k) \tag{32}$$

$$\mathcal{J}(x_{t|t}^{k+1}, p_t^{k+1} q_t^{k+1}) \geq \mathcal{J}(x_{t|t}^k, p_t^{k+1} q_t^k) \tag{33}$$

Combining (32) and (33), we conclude that the sequence $\mathcal{J}(x_{t|t}^k, p_t^k, q_t^k)$ produced by Algorithm 1 is non-decreasing. Furthermore, the objective function (12) is bounded above. According to the monotone convergence theorem [48], we can draw the conclusion that (31) holds.

To sum up, our proposed robust filtering algorithm yields the same solution of the original problem through the half-quadratic technique and its iteratively generated sequence converges. The proof of convergence is now complete.

IV. ROBUST NONLINEAR KALMAN SMOOTHER WITH MCC

In this section, we derive the MCC based robust Kalman smoother along with the similar clue in Section III. In the Bayesian smoothing paradigm, the posterior density of the states $p(x_{0:T}|y_{1:T})$ is given by

$$p(x_{0:T}|y_{1:T}) \propto p(x_0) \prod_{t=1}^{T} p(x_t|x_{t-1}) \prod_{t=1}^{T} p(y_t|x_t) \tag{34}$$

The MAP estimate of the smoothed states is obtained by minimizing the negative log posterior density $p(x_{0:T}|y_{1:T})$, i.e.,

$$\hat{x}_{0:T|T} = \arg\min_{x_{0:T}} \left(-\log(p(x_0)) - \sum_{t=1}^{T} \log(p(x_t|x_{t-1})) \right.$$
$$\left. - \sum_{t=1}^{T} \log(p(y_t|x_t)) \right) \tag{35}$$

Under the Gaussian assumption and dropping the terms that do not relate on $x_{0:T}$, $\hat{x}_{0:T|T}$ can be obtained by

$$\hat{x}_{0:T|T} = \arg\min_{x_{0:T}} \left(\frac{1}{2} \sum_{t=1}^{T} \|x_t - f(x_{t-1})\|_{Q_{t-1}^{-1}}^2 \right.$$
$$\left. + \frac{1}{2} \sum_{t=1}^{T} \|y_t - h(x_t)\|_{R_t^{-1}}^2 + \frac{1}{2} \|x_0 - \hat{x}_{0|0}\|_{P_{0|0}^{-1}}^2 \right) \tag{36}$$

Similarly, (36) can be written in terms of fitting errors as

$$\hat{x}_{0:T|T} = \arg\min_{x_{0:T}} \left(\frac{1}{2} \sum_{t=0}^{T} \alpha_t^T \alpha_t + \frac{1}{2} \sum_{t=1}^{T} \beta_t^T \beta_t \right)$$
$$= \arg\min_{x_{0:T}} \left(\frac{1}{2} \sum_{t=0}^{T} \sum_{i=1}^{n} \alpha_{t,i}^2 + \frac{1}{2} \sum_{t=1}^{T} \sum_{j=1}^{m} \beta_{t,j}^2 \right) \tag{37}$$

where β_t has a same definition in (8), α_t is redefined as

$$\alpha_t = \begin{cases} P_{0|0}^{-1/2} (x_0 - \hat{x}_{0|0}), & t = 0 \\ Q_{t-1}^{-1/2} (x_t - h(x_{t-1})), & t \neq 0 \end{cases} \tag{38}$$

Akin to the robust nonlinear Kalman filter, we replace the quadratic loss with the weighted instantaneous correntropy for each component of the normalized error, resulting in the following objective function for the nonlinear robust smoother

$$\hat{x}_{0:T|T} = \arg\max_{x_{0:T}} \left(\sum_{t=0}^{T} \sum_{i=1}^{n} a_{t,i} \kappa_{\sigma_{t,i}}(\alpha_{t,i}) \right.$$
$$\left. + \sum_{t=1}^{T} \sum_{j=1}^{m} b_{t,j} \kappa_{\eta_{t,j}}(\beta_{t,j}) \right) \tag{39}$$

where $\{a_{t,i}, b_{t,j}\}$ are parameters that should be selected to maintain the consistency of the Gaussian approximation smoother when the SSM does not involve the impulsive noises, similarly to the filtering case discussed in Section III-A. According to Proposition 1, (39) can be transformed as

$$\hat{x}_{0:T|T} = \arg\max_{x,p,q} \left(\sum_{t=0}^{T} \sum_{i=1}^{n} a_{t,i} \left(\frac{p_{t,i}}{2\sigma_{t,i}^2} \alpha_{t,i}^2 - \psi(p_{t,i}) \right) \right.$$
$$\left. + \sum_{t=1}^{T} \sum_{j=1}^{m} b_{t,i} \left(\frac{q_{t,i}}{2\eta_{t,i}^2} \beta_{t,i}^2 - \psi(q_{t,i}) \right) \right) \quad (40)$$

where $x = \{x_t\}_{t=0}^T$, $p = \{p_t\}_{t=0}^T$ and $q = \{q_t\}_{t=1}^T$ are the collections of all variables to be optimized.

Again, the augmented objective function can be solved in an iterative manner via the half-quadratic method. First of all, fixing the states and optimizing (40) with respect to p and q according to Proposition 1, we obtain the same updating equations for $\{p_{t,i}, q_{t,j}\}$ as shown in (13)(14). Next, keeping p and q fixed and dropping the items that do not relate to the state, one can update the smoothed state $\hat{x}_{0:T|T}$ though

$$\hat{x}_{0:T|T} = \arg\max_{x_{0:T}} \left(\sum_{t=0}^{T} \sum_{i=1}^{n} \frac{p_{t,i} a_{t,i}}{2\sigma_{t,i}^2} \alpha_{t,i}^2 \right.$$
$$\left. + \sum_{t=1}^{T} \sum_{j=1}^{m} \frac{q_{t,i} b_{t,i}}{2\eta_{t,i}^2} \beta_{t,i}^2 \right) \quad (41)$$

Similarly, to maintain the consistency of the Gaussian approximation smoother when the SSM dose not involve any impulsive noise, the coefficients should be selected by (18) and (19). Substituting (18) and (19) into (41), the objective function is updated as

$$\hat{x}_{0:T|T} = \arg\max_{x_{0:T}} \left(\sum_{t=0}^{T} \sum_{i=1}^{n} \frac{1}{2} p_{t,i} \alpha_{t,i}^2 \right.$$
$$\left. + \sum_{t=1}^{T} \sum_{j=1}^{m} \frac{1}{2} q_{t,i} \beta_{t,i}^2 \right) \quad (42)$$

Writing in matrix format, we get

$$\hat{x}_{0:T|T} = \arg\min_{x_{0:T}} \left(\frac{1}{2} \sum_{t=0}^{T} \alpha_t^T \Psi_t \alpha_t + \frac{1}{2} \sum_{t=1}^{T} \beta_t^T \Phi_t \beta_t \right) \quad (43)$$

where $\Psi_t \in \mathbb{R}^{n \times n}$ and $\Phi_t \in \mathbb{R}^{m \times m}$ are diagonal matrixes, given by (22) and (23), respectively. Furthermore, substituting the expression of α_t and β_t, namely (38) and (8), into (43), we obtain

$$\hat{x}_{0:T|T} = \arg\min_{x_{0:T}} \left(\frac{1}{2} \sum_{t=1}^{T} \|x_t - f(x_{t-1})\|_{\bar{Q}_{t-1}^{-1}}^2 \right.$$
$$\left. + \frac{1}{2} \sum_{t=1}^{T} \|y_t - h(x_t)\|_{\bar{R}_t^{-1}}^2 + \frac{1}{2} \|x_0 - \hat{x}_{0|0}\|_{\bar{P}_{0|0}^{-1}}^2 \right) \quad (44)$$

where

$$\bar{Q}_{t-1} = Q_{t-1}^{1/2} \Psi_t^{-1} Q_{t-1}^{T/2} \quad (45)$$
$$\bar{R}_t = R_t^{1/2} \Phi_t^{-1} R_t^{T/2} \quad (46)$$
$$\bar{P}_{0|0} = P_{0|0}^{1/2} \Psi_0^{-1} P_{0|0}^{T/2} \quad (47)$$

The objective function (44) can be solved by CKS, which provides updates of the state $\hat{x}_{0:T|T}$. The iterative procedure is repeated till some convergence criterion is met. The detailed robust cubature Kalman smoother (RCKS) is summarized in Algorithm 2.

Algorithm 2 MCC Based Robust Cubature Kalman Smoother

Input: $y_{1:T}, \hat{x}_0, P_0, Q_{1:T}, R_{1:T}$.
Output: $\hat{x}_{t|T}$ and $P_{t|T}$ for $t = 1 : T$.
Initialized: $k = 0$, $\Psi_t = I_n$, $\Phi_t = I_m$.
repeat
    1. Update $\bar{Q}_{t-1}, \bar{R}_t$ and $\bar{P}_0$ via (45)-(47);
    2. Update $\hat{x}_{0:T|T}^k$ and $P_{0:T|T}^k$ via CKS;
    3. Update Φ_t and Ψ_t via (38), (8), (13), (14), (22) and (23);
    4. $k = k + 1$;
until a convergence criterion is met;
$\hat{x}_{t|T} = \hat{x}_{t|T}^{(k-1)}$ and $P_{t|T} = P_{t|T}^{(k-1)}$;

It can be seen that the proposed robust nonlinear Kalman smoother actually contains two steps in each iteration, i.e., updating the noise covariance and running the conventional Gaussian approximation smoother. One can follow similar steps used in Section III-C and show that the proposed nonlinear robust Kalman smoother converges.

V. NUMERICAL SIMULATIONS

Numerical examples are presented in this section to illustrate the performance of the proposed RCKF and RCKS in the presence of heavy-tailed noises. We also compare the state estimates of the proposed algorithms with those of the conventional CKF [4], conventional CKS [43], linear regression and MCC based robust Kalman filter (LRKF) [27], nonlinear regression based Huber Kalman filter (HRKF) [16] and variational Bayesian based student's cubature Kalman smoother (TCKS) [21]. The LRKF is originally proposed in the unscented Kalman filter (UKF) framework and here for consistency with the other compared algorithms, we set parameters $(\alpha, \beta, \kappa) = (1, 0, 0)$ in LRKF, which makes it functionally equivalent to a CKF. Other design parameters in the aforementioned algorithms are set as recommended in their literatures. In the simulation, we take the following convergence criterion, $\|\hat{x}_{t|t}^k - \hat{x}_{t|t}^{k-1}\|/\|\hat{x}_{t|t}^{k-1}\| \leq 10^{-6}$ in filtering and $\max(\|\hat{x}_{t|T}^k - \hat{x}_{t|T}^{k-1}\|/\|\hat{x}_{t|T}^{k-1}\|) \leq 10^{-6}$ for $t = 1, \cdots, T$ in smoothing. For clarity, the algorithms studied in this section are summarized as follows:

NOMENCLATURE

CKF	cubature Kalman filter [4]
CKS	cubature Kalman smoother [43]
RCKF	proposed MCC based filter
RCKS	proposed MCC based smoother
LRKF	linear regression and MCC based Kalman filter [27]
HRKD	nonlinear regression based Huber Kalman filter [16]
TCKS	variational Bayesian based student's t cubature Kalman smoother [21]

A. VAN DER POL OSCILLATOR

The Van der Pol oscillator (VPO) model [49] is usually used as a benchmark for testing the performance of nonlinear filters and smoothers. The continuous-time nonlinear dynamics of the VPO is governed by the following differential equations

$$\dot{x}_1 = x_2 \tag{48}$$

$$\dot{x}_2 = \mu \left(1 - x_1^2\right) x_2 - x_1 \tag{49}$$

where μ is a scalar parameter indicating the nonlinearity. Discretization of the VPO yields

$$\boldsymbol{x}_{t+1} = \begin{pmatrix} x_{1,t} + \int_t^{t+\delta} x_2 \, dt \\ x_{2,t} + \int_t^{t+\delta} \left(\mu \left(1 - x_1^2\right) x_2 - x_1\right) dt \end{pmatrix} + \boldsymbol{\omega}_t \tag{50}$$

where $\boldsymbol{x}_t = [x_{1,t}, x_{2,t}]^T$ is the state, $\boldsymbol{\omega}_t$ is the process noise to model the discretization error, and δ is the sampling interval. The integral terms in (50) are numerically calculated by the fourth-order Runge-Kutta scheme with a time step of the sampling interval. Furthermore, we assume that the state is measured by the following noise disturbed nonlinear procedure

$$y_t = (x_{1,t} - 1)^2 + 1 + v_t \tag{51}$$

Both noises are generated from the mixed-Gaussian model [19]:

$$\boldsymbol{\omega}_{t-1} \sim (1 - p_1)\mathcal{N}(0, \boldsymbol{Q}_{t-1}) + p_1 \mathcal{N}(0, \phi_1 \boldsymbol{Q}_{t-1}) \tag{52}$$

$$v_t \sim (1 - p_2)\mathcal{N}(0, R_t) + p_2 \mathcal{N}(0, \phi_2 R_t) \tag{53}$$

where p_1 and p_2 are the contaminating parameters, ϕ_1 and ϕ_2 are the scaling factors indicating the strength of the contaminating noises, and $\boldsymbol{Q}_{t-1}$ and R_t are the covariance matrices of the nominal process noise and measurement noise, respectively.

In the simulation, we select $\mu = 1$ and total samples $T = 120$ with the sampling interval $\delta = 0.1$s. The initial state $\hat{\boldsymbol{x}}_{0|0}$ follows $\mathcal{N}(\boldsymbol{x}_0, 0.01\boldsymbol{I}_2)$ with $\boldsymbol{x}_0 = [0, -0.5]^T$. The covariance of the noises are set as $\boldsymbol{Q}_{t-1} = 0.01\boldsymbol{I}_2$ and $R_t = 1$, respectively. The covariance of the contaminating part in each noise is controlled by $\phi_1 = 10$ and $\phi_2 = 50$. Specifically, three scenarios are considered:
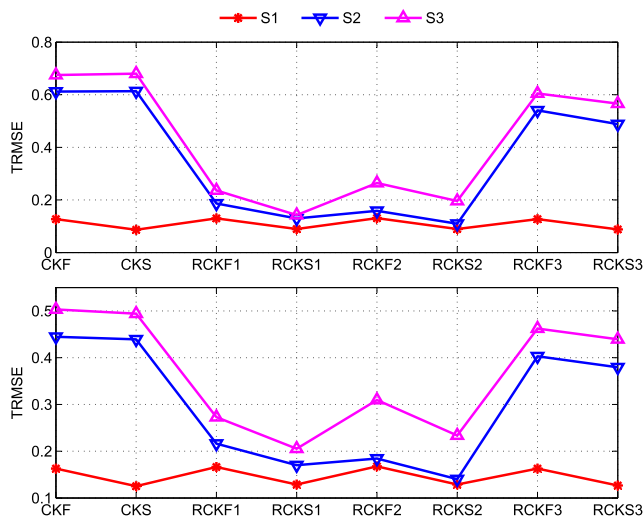
FIGURE 1. The TRMSEs of different methods for x_1 (top) and x_2 (bottom).
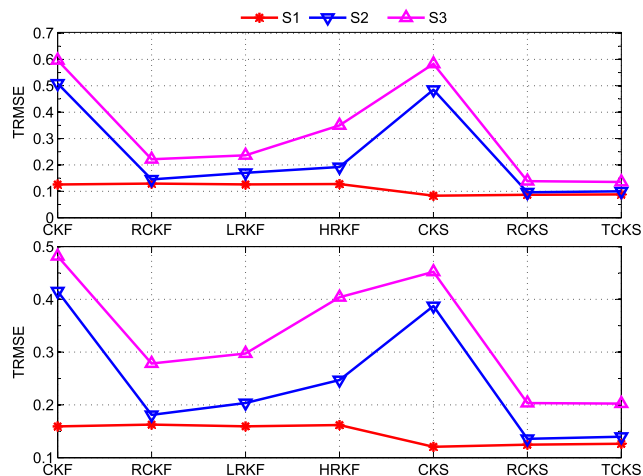
**FIGURE 2.** The TRMSEs of the different methods for x_1 (top) and x_2 (bottom).

- Scenario 1 (S1): all noises are Gaussian, i.e., $p_1 = 0$ and $p_2 = 0$;
- Scenario 2 (S2): the process noise is a Gaussian noise while the measurement noise is a heavy-tailed noise, i.e., $p_1 = 0$ and $p_2 = 0.2$;
- Scenario 3 (S3): both process and measurement noises are heavy-tailed noises, i.e., $p_1 = 0.2$ and $p_2 = 0.2$.

For fair comparison, we implement $L = 1000$ Monte Carol runs in each scenario. We use the implementation time (IT) and time-averaged root mean square error (TRMSE) as performance metrics. The TRMSE is defined as

$$\mathrm{TRMSE} = \frac{1}{T} \sum_{t=1}^{T} \sqrt{\frac{1}{L} \sum_{i=1}^{L} \left(x_{1,t}^i - \hat{x}_{1,t}^i\right)^2} \tag{54}$$

where $x_{1,t}^i$ and $\hat{x}_{1,t}^i$, respectively, denote the first component of the true and estimated state at time t in the i-th Monte Carlo run. In addition, the TRMSE for the second component can be similarly defined.
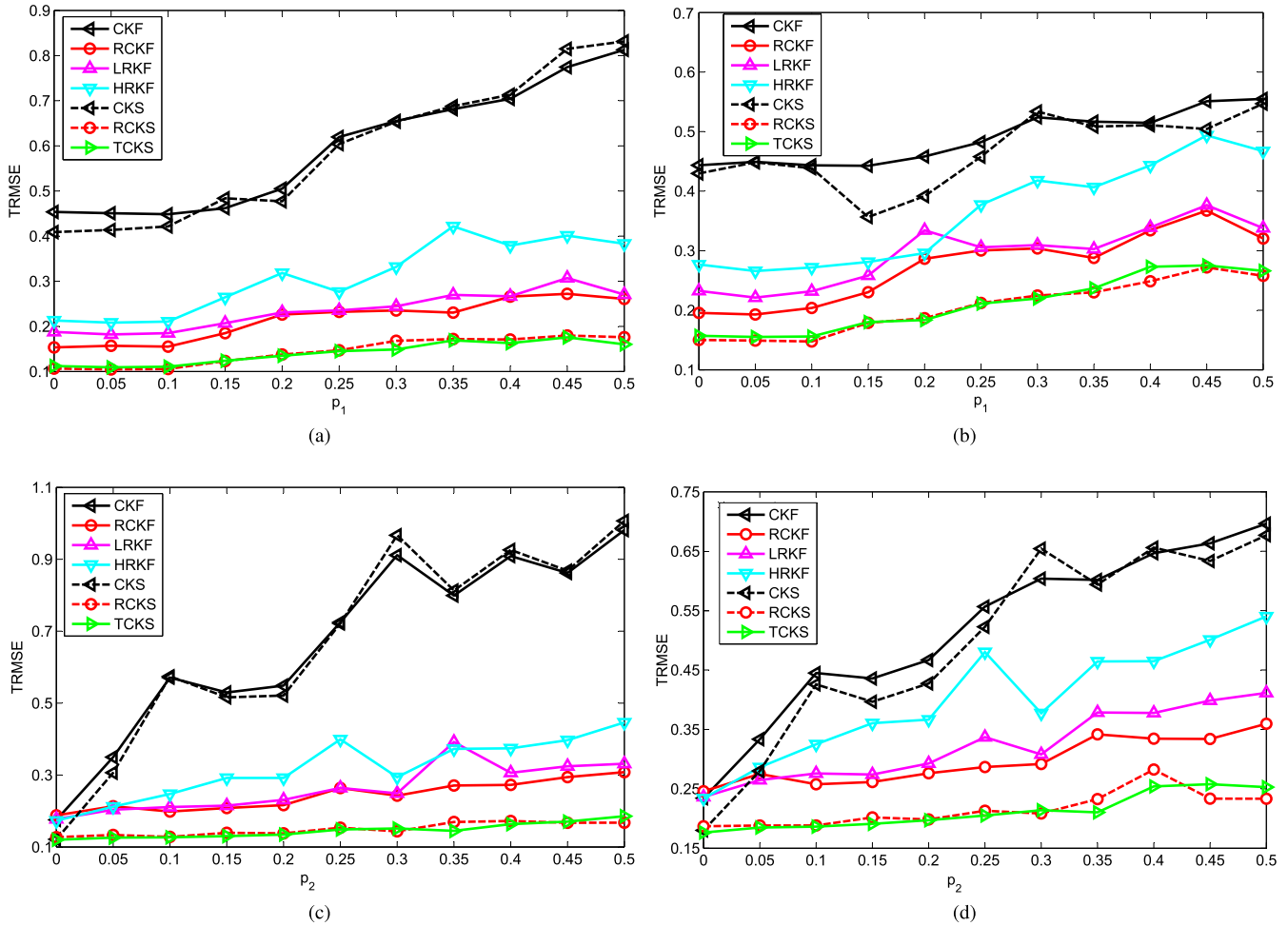
FIGURE 3. TRMSE of different estimators via the contamination ratio. Top: varying p_1 with fixed $p_2 = 0.2$, (a) TRMSE of x_1; (b) TRMSE of x_2. Bottom: varying p_2 with fixed $p_1 = 0.2$, (c) TRMSE of x_1; (d) TRMSE of x_2.

The kernel size is the only design parameter in the proposed algorithms and it has a significant impact on the performance of the RCKF and RCKS. Therefore, we studied the performance of the proposed algorithms with three different kernel sizes. The labels of the resulting methods and their associated kernel sizes are shown in Table 1. The conventional CKF and CKS are also implemented as a performance benchmark.

TABLE 1. The kernel sizes of different methods.

Methods	Kernel size
RCKF1/RCKS1	$\sigma_t = 2, \ \eta_t = 2$
RCKF2/RCKS2	$\sigma_t = 20, \ \eta_t = 2$
RCKF3/RCKS3	$\sigma_t = 20, \ \eta_t = 20$

Fig. 1 shows the TRMSEs while the IT is collected in Table 2. It can be seen, as expected, the smoothers have higher accuracies at the cost of more computational time than their filtering counterparts due to the use of backward procedure. The IT of the robust estimators with different kernel sizes in three scenarios are similar, which implies the IT of the proposed algorithms is insensitive to the kernel size.

In the S1, the robust algorithms are similar to the conventional ones, which suggest that the robust algorithms are

TABLE 2. The ITs of different methods.

Method	IT/s			Method	IT/s		
	S1	S2	S3		S1	S2	S3
CKF	37.1	36.8	36.6	RCKF1	142.6	140.9	130.8
CKS	44.5	43.8	42.1	RCKS1	643.0	590.7	533.3
RCKF2	136.7	133.0	122.4	RCKF3	127.8	130.8	122.0
RCKS2	648.5	588.6	532.4	RCKS3	642.8	589.8	531.9

consistent with the conventional ones. When heavy-tailed noises are involved, taking their characteristics into account in general pays off in terms of accuracy, while the computational time increase because of their iterative procedure. It is observed that RCKF3/RCKS3, which employ a larger kernel size, perform the worst among the robust algorithms. RCKF1/RCKS1 and RCKF2/RCKS2 are slightly different in their perfermance, which is caused by their different kernel sizes.

Next we present comparison results for different algorithms. For fairness, we set all kernel sizes to 2 in LRKF, RCKF and RCKS. Fig. 2 shows the TRMSEs of the different algorithms in the three scenarios and the IT of the different algorithms are given in Table 3. It is clear that

all robust algorithms have a similar performance as that of the conventional estimators in the scenario when all noises are Gaussian. However, in the presence of the heavy-tailed noises, the robust estimators have a significant improvement over the conventional algorithms, which does not account for the presence of outliers. It is observed that the proposed RCKF outperforms HRKF. This is mainly because the MCC is instinctively more suitable for heavy-tailed noise than Huber's function. It is also seen that RCKF has a slight gain than LRKF. This is because, while both algorithms are derived from MCC, LRKF is based on a linear regression which introduces the linearzation error. Although RCKF is better than LRKF and HRKF in accuracy, the former has a higher IT than the latter, by 30% and 95%, respectively. On the other hand, all three smoothers have smaller TRMSEs than their filtering counterparts. The proposed RCKS has a similar performance to that of TCKS and both outperform the traditional CKS. TCKS, however, needs about 40% more IT than the proposed RCKS.

The TRMSE versus the increase of the contamination ratio is also studied, as shown in Fig. 3. The TRMSEs of the two traditional algorithms rise considerably, while the inverse is true for all robust algorithms except HRKF. The slightly increasing TRMSEs of the robust algorithms suggests that these algorithms are relatively insensitive to the level of the heavy-tailed noise. The TRMSEs of HRKF have a slight increase when the contamination ratio is relatively small and then the increase is more notable when the contamination ratio becomes larger, confirming that Huber's function based robust estimators performs well only at relativly low levels of contamination. We have also compared with the EKF and UKF, and found the latter two are similar to the CKF for the considered VPO problems. The results of the EKF and UKF are not included in Fig. 3 for clarity.

B. AGILE TARGET TRACKING
Here we consider the problem of tracking an agile target which is observed by an active radar located at the original point. The outliers may occur in both the process and measurement procedure due to the rapid motion of the target or disturbances experienced by the radar. The SSM is formulated as [21] and [50]–[52]

$$x_t = \begin{bmatrix} I_2 & \delta I_2 \\ 0 & I_2 \end{bmatrix} x_{t-1} + \omega_{t-1} \quad (55)$$

$$y_t = \begin{bmatrix} \sqrt{a_t^2 + b_t^2} \\ \mathrm{atan2}(b_t, a_t) \end{bmatrix} + v_t \quad (56)$$

TABLE 3. The implementation time of different methods.

Method	IT/s			Method	IT/s		
	S1	S2	S3		S1	S2	S3
CKF	37.1	36.8	36.6	CKS	44.5	43.8	42.1
RCKF	136.7	133.0	132.4	RCKS	648.5	588.6	532.4
LRKF	95.1	98.3	90.4	TCKS	1097.2	964.4	941.1
HRKF	66.1	68.1	63.9				
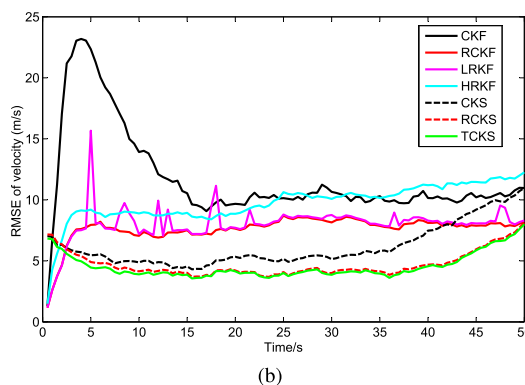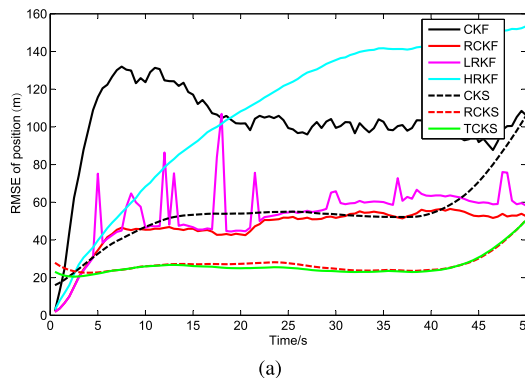
FIGURE 4. The RMSE of different estimators when $p_1 = 0.2$ and $p_2 = 0.2$: (a) RMSE of position; (b) RMSE of velocity.

where $x_t = [a_t, b_t, \dot{a}_t, \dot{b}_t]^T$ is the state; a_t, b_t, $\dot{a}_t$ and $\dot{b}_t$ denote the positions and corresponding velocities in Cartesian coordinates; $\delta = 0.5$s is the sampling interval and atan2 is the four-quadrant inverse tangent function. In the simulation, the initial state $\hat{x}_{0|0}$ is randomly chosen from $\mathcal{N}(x_0, P_0)$ with the true initial state $x_0 = [-10000\text{m}, 10000\text{m}, 30\text{m/s}, -40\text{m/s}]^T$ and corresponding error covariance $P_0 = 100I_4$. Outlier-corrupted process and measurement noises are generated from the following Gaussian mixture model:

$$\omega_{t-1} \sim 0.8\mathcal{N}(0, Q_{t-1}) + 0.2\mathcal{N}(0, 10Q_{t-1}) \quad (57)$$

$$v_t \sim 0.8\mathcal{N}(0, R_t) + 0.2\mathcal{N}(0, 50R_t) \quad (58)$$

where the nominal covariance matrices for the process and measurement noise, i.e. Q and R, are given as

$$Q_{t-1} = \begin{bmatrix} \frac{\delta^3}{3}I_2 & \frac{\delta^2}{2}I_2 \\ \frac{\delta^2}{2}I_2 & \delta I_2 \end{bmatrix}, \quad R_t = \begin{bmatrix} 100\text{m}^2 & 0 \\ 0 & 16(\text{mrad})^2 \end{bmatrix} \quad (59)$$

$L = 1000$ independent Monte-Carlo runs are implemented and in each run $T = 200$ noisy measurements are collected. The root mean square error (RMSE) of position and velocity are performed as performance metrics, which are defined as

$$\mathrm{RMSE}_{\text{pos}}(t) = \sqrt{\frac{1}{L}\sum_{i=1}^{L}\left((a_t^i - \hat{a}_t^i)^2 + (b_t^i - \hat{b}_t^i)^2\right)} \quad (60)$$

where a_t^i and $\hat{a}_t^i$ have similar meanings in (54). Akin to position, the RMSE of velocity can also be formulated.

Fig. 4 shows the RMSEs of the position and velocity for the compared algorithms. It can be seen that the robust filters have a significant improvement except HRKF. Although both RCKF and LRKF are based on MCC, the proposed RCKF performs better than the LRKF, mainly because LRKF is based a linear regression, which introduces the linearzation error. The superiority of both RCKS and TCKS are also illustrated in Fig. 4 where RCKS has a similar performance as the TCKS.

VI. CONCLUSION

A new unified framework for robust Kalman filtering and smoothing based on the MCC has been proposed in this work. To ease the understanding of the proposed approach, we integrate the conventional cubature Kalman filter and smoother into our framework. The proposed robust Kalman filter and smoother is derived under the MCC, instead of the well-known MMSE, to cope with heavy-tailed noises in both the process and observation procedures. The intractable objective function based on MCC is then solved by the half-quadratic optimization method in an iterative manner. At each iteration the problem is reduced to a quadratic optimization problem, which can be efficiently solved by the conventional nonlinear Kalman filter and smoother. Numerical simulations demonstrate the superiority of the proposed robust filter and smoother when dealing with outliers in both states and measurements. The simulation results show that the proposed algorithms outperform state-of-art methods at comparable computational burden. One interesting future research direction is to investigate how to choose the kernel bandwidth of the MCC.

APPENDIX A
CUBATURE KALMAN FILTER AND SMOOTHER

We consider the SSM described in (1) and (2) without the contaminated noise (i.e., $\rho_1 = \rho_2 = 0$) and the process and measurement noise follow

$$v_{t-1} \sim \mathcal{N}(0, Q_{t-1})$$
$$w_t \sim \mathcal{N}(0, R_t)$$

The CKF has two steps involved, i.e., predicting and filtering. After the CKF procedure, the CKS is implemented by a backward smoothing step. The CKF and CKS are briefly summarized here for easy reference. The algorithms are initialized with $x_0 \sim \mathcal{N}(\hat{x}_{0|0}, P_{0|0})$ and the basic weighted cubature point set is given by $\{\xi_i, \omega_i\}$ for $i = 1, \cdots, 2n$, where $\xi_i = \sqrt{n}[I]_i$, $[I] = [I_n, -I_n]$ and $\omega_i = 1/(2n)$.

Prediction:

$$\chi_{i,t-1} = f(\xi_{i,t-1}) \tag{A.1}$$

$$\hat{x}_{t|t-1} = \sum_{i=1}^{2n} \omega_i \chi_{i,t-1} \tag{A.2}$$

$$P_{t|t-1} = \sum_{i=1}^{2n} \omega_i (\chi_{i,t-1} - \hat{x}_{t|t-1}) \\ \times (\chi_{i,t-1} - \hat{x}_{t|t-1})^T + Q_{t-1} \tag{A.3}$$

where $\xi_{i,t-1}$ is the transformed sigma point related to distribution $\mathcal{N}(\hat{x}_{t-1|t-1}, P_{t-1|t-1})$, i.e.,

$$P_{t-1|t-1} = S_{t-1|t-1}S_{t-1|t-1}^T$$
$$\xi_{i,t-1} = S_{t-1|t-1}\xi_i + \hat{x}_{t-1|t-1}$$

Filtering:

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - \hat{y}_t) \tag{A.4}$$

$$P_{t|t} = P_{t|t-1} - K_t P_{yy} K_t^T \tag{A.5}$$

$$K_t = P_{xy} P_{yy}^{-1} \tag{A.6}$$

where

$$\psi_{i,t} = h(\pi_{i,t}), \quad \hat{y}_t = \sum_{i=1}^{2n} \omega_i \psi_{i,t} \tag{A.7}$$

$$P_{yy} = \sum_{i=1}^{2n} \omega_i (\psi_{i,t} - \hat{y}_t)(\psi_{i,t} - \hat{y}_t)^T + R_t \tag{A.8}$$

$$P_{xy} = \sum_{i=1}^{2n} \omega_i (\chi_{i,t} - \hat{x}_{t|t-1})(\psi_{i,t} - \hat{y}_t)^T \tag{A.9}$$

where $\pi_{i,t}$ is the transformed sigma point related to distribution $\mathcal{N}(\hat{x}_{t|t-1}, P_{t|t-1})$, i.e.,

$$P_{t|t-1} = S_{t|t-1}S_{t|t-1}^T, \quad \pi_{i,t} = S_{t|t-1}\xi_i + \hat{x}_{t|t-1}$$

Smoothing

$$\hat{x}_{t|T} = \hat{x}_{t|t} + D_{t+1}(\hat{x}_{t+1|T} - \hat{x}_{t+1|t}) \tag{A.10}$$

$$P_{t|T} = P_{t|t} + D_{t+1}(P_{t+1|T} - P_{t+1|t})D_{t+1}^T \tag{A.11}$$

where

$$D_{t+1} = C_{t+1}P_{t+1|t}^{-1} \tag{A.12}$$

$$C_{t+1} = \sum_{i=1}^{2n} \omega_i(\xi_{i,t} - \hat{x}_{t|t})(\chi_{i,t} - \hat{x}_{t+1|t})^T \tag{A.13}$$

Here we should note that the filtered state is the same as the smoothed state at the end instant, so does their corresponding error covariance.

APPENDIX B
CORRENTROPY AND MAXIMUM CORRENTROPY CRITERION

The correntropy, which is first proposed in [26], is a concept in information theoretic learning to deal with non-Gaussian noise. The correntropy is a generalized similarity measure between two arbitrary scalar random variables X and Y, defined by

$$V(X, Y) = E(\kappa_\sigma(X - Y)) \tag{B.1}$$

where $E(\cdot)$ is the expectation and $\kappa(\cdot)$ is a kernel function with a control parameter σ, which satisfies Mercer's Theorem [45]. The commonly used kernel, the Gaussian kernel, is considered in this paper, i.e.,

$$\kappa_\sigma(X - Y) = \exp(-\frac{e^2}{2\sigma^2}) \tag{B.2}$$

where $e = X - Y$ is the difference between two variables, and $\sigma > 0$ is the bandwidth of the Gaussian kernel.

Calculating the exact value of $V(X, Y)$ requires the joint distribution of X and Y which is barely known in practice. However, finite number of pairwise samples $\{x_i, y_i\}_{i=1}^N$ are often available. The correntropy can be estimated by

$$V(x, y) \approx \frac{1}{N} \sum_{i=1}^N \kappa(x_i - y_i). \tag{B.3}$$

Correntropy has a number of nice properties that make it useful for non-Gaussian signal processing, especially in the impulsive noise environment. Correntropy is symmetric, positive and bounded. In addition, unlike the global similarity measure-mean square error (MSE) which only contains the second-order statistics, correntropy incorporates all even order moments [26]. In geometric meaning, MSE gives the ℓ_2 norm distance over $\{x, y\}$ while correntropy offers a hybird norm distance. Specifically, when two points are similar, correntropy behaves like the ℓ_2 norm distance while it approaches the ℓ_1 norm distance with increasing difference between two points. Finally, correntropy is equivalent to the ℓ_0 norm distance if two points are far apart.

Based on the geometric meaning of correntropy, maximizing the correntropy of two different random variables can be used as a criterion in dealing with non-Gaussian noise problem, especially heavy-tailed noises, which leads to MCC. MCC has a close relationship with the M-estimator [26], and it is actually equivalent to the Welsch M-estimator [29]. The superior performance of MCC in handling with the impulse noise has also been reported in [29]. Hence MCC is a promising option to design robust filters and smoothers.

APPENDIX C
PROOF OF PROPOSITION 1
In order to prove Proposition 1, we first introduce the following theorem from [53]:

Theorem 1: Let $f(x)$ be a function that satisfies the following conditions:

1) $f(x) \geq 0 \ \forall x$ with $f(0) = 0$;
2) $f(x) = f(-x)$;
3) $f(x)$ continuously differentiable;
4) $f'(x) \geq 0 \ \forall x \geq 0$;
5) $f'(x)/(2x)$ continuous and strictly decreasing on $[0, +\infty)$;
6) $\lim_{x \to +\infty} (f'(x)/(2x)) = 0$;
7) $\lim_{x \to 0^+} (f'(x)/(2x)) = M$ where $0 < M < +\infty$

then

1) there exists a strictly convex and decreasing function $g : (0, M] \to [0, \beta)$ where

$$\beta = \lim_{x \to +\infty} \left(f(x) - x^2 \frac{f'(x)}{2x} \right) \tag{C.1}$$

such that

$$f(x) = \inf_{0 < w \leq M} \left(wx^2 + g(w) \right) \tag{C.2}$$

2) for every fixed $\tilde{x}$, the value w^m for which the minimum is reached, i.e., such that

$$\inf_{0 < w \leq M} \left(wx^2 + g(w) \right) = \left(w^m \tilde{x}^2 + g(w^m) \right) \tag{C.3}$$

is unique and given by

$$w^m = \frac{f'(\tilde{x})}{2\tilde{x}} \tag{C.4}$$

To use the above theorem to prove Proposition 1, let $f(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}$. It is easy to check that $f(x)$ satisfies 1) to 7) in Theorem 1. Therefore, we conclude that there exists a convex function $\theta(w) : (0, M] \to [0, \beta)$ such that

$$1 - e^{-\frac{x^2}{2\sigma^2}} = \inf_{0 < w \leq M} \left(wx^2 + \theta(w) \right) \tag{C.5}$$

with $M = 1/(2\sigma^2)$, $\beta = 1$, and for a fixed x, the minimum is reached at $w = \frac{1}{2\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$.

We can equivalently rewrite (C.5) in terms of the supremum as

$$e^{-\frac{x^2}{2\sigma^2}} = \sup_{0 < w \leq M} \left(-wx^2 - \theta(w) + 1 \right) \tag{C.6}$$

Define $p = -2w\sigma^2$, and then (C.6) is transformed as

$$e^{-\frac{x^2}{2\sigma^2}} = \sup_{-1 \leq p < 0} \left(\frac{p}{2\sigma^2} x^2 - \theta(-\frac{p}{2\sigma^2}) + 1 \right) \tag{C.7}$$

Furthermore, define $\psi(p) = \left(\theta(-\frac{p}{2\sigma^2}) - 1 \right) : [-1, 0) \to [-1, 0)$. We obtain

$$\kappa_\sigma(x) = \sup_{-1 \leq p < 0} \left(\frac{p}{2\sigma^2} x^2 - \psi(p) \right) \tag{C.8}$$

and for a fixed x the supremem is reached at

$$p = -2\sigma^2 \times \frac{1}{2\sigma^2} e^{-\frac{x^2}{2\sigma^2}} = -\kappa_\sigma(x) \tag{C.9}$$

Since $\psi(p)$ is obtained by an affine transformation from $\theta(w)$, $\psi(p)$ is a convex function due to the convexity invariance under affine maps.

REFERENCES
[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, D, J. Basic Eng.*, vol. 82, pp. 35–45, 1960.
[2] B. D. Anderson and J. B. Moore, *Optimal Filtering*. New York, NY, USA: Prentice-Hall, 1979.
[3] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.
[4] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.
[5] F. Gustafsson and G. Hendeby, "Some relations between extended and unscented Kalman filters," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 545–555, Feb. 2012.
[6] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2910–2921, Aug. 2006.
[7] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, 1965.
[8] S. Sarkka and J. Hartikainen, "On Gaussian optimal smoothing of nonlinear state space models," *IEEE Trans. Autom. Control*, vol. 55, no. 8, pp. 1938–1941, Aug. 2010.

[9] I. Bilik and J. Tabrikian, "Maneuvering target tracking in the presence of glint using the nonlinear Gaussian mixture Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 246–262, Jan. 2010.

[10] I. C. Schick and S. K. Mitter, "Robust recursive estimation in the presence of heavy-tailed observation noise," *Ann. Statist.*, vol. 22, no. 2, pp. 1045–1080, 1994.

[11] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.

[12] 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.

[13] C. Masreliez and R. Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *IEEE Trans. Autom. Control*, vol. 22, no. 3, pp. 361–371, Jun. 1977.

[14] P. Huber and E. Ronchetti, *Robust Statistics*. New York, NY, USA: Wiley, 2009.

[15] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.

[16] C. D. Karlgaard, "Nonlinear regression Huber–Kalman filtering and fixed-interval smoothing," *J. Guid., Control, Dyn.*, vol. 38, no. 2, pp. 322–330, 2015.

[17] L. Chang and K. Li, "Unified form for the robust Gaussian information filtering based on M-estimate," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 412–416, Apr. 2017.

[18] J.-S. Hu and C.-H. Yang, "Second-order extended H_∞ filter for nonlinear discrete-time systems using quadratic error matrix approximation," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3110–3119, Jul. 2011.

[19] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "Robust student's t based nonlinear filter and smoother," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2586–2596, Oct. 2016.

[20] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.

[21] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 468–472, Apr. 2016.

[22] A. Y. Aravkin, B. M. Bell, J. V. Burke, and G. Pillonetto, "An ℓ_1-Laplace robust Kalman smoother," *IEEE Trans. Autom. Control*, vol. 56, no. 12, pp. 2898–2911, Dec. 2011.

[23] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Robust and trend-following Student's t Kalman smoothers," *SIAM J. Control Optim.*, vol. 52, no. 5, pp. 2891–2916, 2014.

[24] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Sparse/robust estimation and Kalman smoothing with nonsmooth log-concave densities: Modeling, computation, and theory," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 2689–2728, 2013.

[25] 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.

[26] 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.

[27] 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.

[28] J. E. Dennis, Jr., and R. E. Welsch, "Techniques for nonlinear least squares and robust regression," *Commun. Statist.-Simul. Comput.*, vol. 7, no. 4, pp. 345–359, 1978.

[29] R. He, W.-S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 261–275, Feb. 2014.

[30] H. Yan, X. Yuan, S. Yan, and J. Yang, "Correntropy based feature selection using binary projection," *Pattern Recognit.*, vol. 44, no. 12, pp. 2834–2842, 2011.

[31] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[32] R. He, B.-G. Hu, W.-S. Zheng, and X.-W. Kong, "Robust principal component analysis based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1485–1494, Jun. 2011.

[33] 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.*, Jul./Aug. 2011, pp. 489–494.

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

[35] 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)*, Mar. 2016, pp. 500–505.

[36] M. V. Kulikova, "Square-root algorithms for maximum correntropy estimation of linear discrete-time systems in presence of non-Gaussian noise," *Syst. Control Lett.*, vol. 108, pp. 8–15, Oct. 2017.

[37] B. Chen, X. Liu, H. Zhao, and J. C. Príncipe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.

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

[39] W. Qin, X. Wang, and N. Cui, "Maximum correntropy sparse Gauss–Hermite quadrature filter and its application in tracking ballistic missile," *IET Radar, Sonar Navigat.*, vol. 11, no. 9, pp. 1388–1396, 2017.

[40] G. Y. Kulikov and M. Kulikova, "Estimation of maneuvering target in the presence of non-Gaussian noise: A coordinated turn case study," *Signal Process.*, vol. 145, pp. 241–257, Apr. 2018.

[41] Y. Wang, W. Zheng, S. Sun, and L. Li, "Robust information filter based on maximum correntropy Criterion," *J. Guid., Control, Dyn.*, vol. 39, no. 5, pp. 1126–1131, 2016.

[42] 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, Dec. 2017.

[43] I. Arasaratnam and S. Haykin, "Cubature Kalman smoothers," *Automatica*, vol. 47, no. 10, pp. 2245–2250, 2011.

[44] 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. 2017.

[45] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

[46] B. Chen, J. Wang, H. Zhao, N. Zheng, and J. C. Príncipe, "Convergence of a fixed-point algorithm under maximum correntropy criterion," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1723–1727, Oct. 2015.

[47] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[48] J. Yeh, *Real Analysis: Theory of Measure and Integration*. Hackensack, NJ, USA: World Scientific, 2006.

[49] R. Kandepu, B. Foss, and L. Imsland, "Applying the unscented Kalman filter for nonlinear state estimation," *J. Process Control*, vol. 18, nos. 7–8, pp. 753–768, 2008.

[50] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.

[51] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation: Theory Algorithms and Software*. Hobken, NJ, USA: Wiley, 2004.

[52] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss–Hermite quadrature," *Proc. IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[53] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, 1997.

HONGWEI WANG (S'18) received the B.S. degree in aerospace engineering from Northwestern Polytechnical University in 2013, where he is currently pursuing the Ph.D. degree with the School of Aeronautics. In 2016, he was a Research Intern with the School of Electrical and Electronic Engineering, Nanyang Technological University. From 2017 to 2018, he was a Visiting Scholar with the Electrical and Computer Engineering Department, Stevens Institute of Technology. His research interests include robust Kalman filtering and smoothing, target tracking, system identification, and decentralized sensors fusion.

HONGBIN LI (M'99–SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1999, all in electrical engineering.

From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. Since 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he became a Professor in 2010. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory in the summers of 2003, 2004, and 2009, respectively. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li has been a member of the IEEE SPS Signal Processing Theory and Methods Technical Committee (TC) and the IEEE SPS Sensor Array and Multichannel TC. He is a member of Tau Beta Pi and Phi Kappa Phi. He received the Sigma Xi Graduate Research Award from the University of Florida in 1999, the Jess H. Davis Memorial Award for excellence in research in 2001 from the Stevens Institute of Technology, the Harvey N. Davis Teaching Award in 2003, the Outstanding Paper Award from the IEEE AFICON Conference in 2011, and the IEEE Jack Neubauer Memorial Award in 2013 from the IEEE Vehicular Technology Society. He has been an Associate Editor for *Signal Processing* (Elsevier), the IEEE Transactions on Signal Processing, the IEEE Signal Processing Letters, and the IEEE Transactions on Wireless Communications, and a Guest Editor for the IEEE Journal of Selected Topics in Signal Processing and the *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities, including serving as a General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, Hoboken, NJ, USA, in 2012.

JUNYI ZUO received the B.E. and M.E. degrees from the School of Aeronautics, Northwestern Polytechnical University, in 2000 and 2003, respectively, and the Ph.D. degree from the School of Automation, Northwestern Polytechnical University, in 2008. He is currently an Associate Professor at the School of Aeronautics, Northwestern Polytechnical University. His research interests include optimal estimation theory and its application, nonlinear filtering, target tracking, and aircraft system identification.

WEI ZHANG was born in 1963. He received the B.S. and M.S. degrees in flight mechanics and the Ph.D. degree in aircraft design from Northwestern Polytechnical University. He was invited as an Associate Professor with the School of Creative Media, City University of Hong Kong, from 2013 to 2014. He is currently a Professor with the School of Aeronautics, Northwestern Polytechnical University. His research interests include aircraft design methodology, system integration and design, scaled flight test, and aircraft system identification.

HEPING WANG was born in 1957. He received the Ph.D. degree from the School of Aeronautics, Northwestern Polytechnical University, in 1995. He is currently a Professor with the School of Aeronautics, Northwestern Polytechnical University. His research interests include flight vehicle design, and aircraft system design. He is a member of the Chinese Society of Aeronautics and Astronautics.

• • •