IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 19, 2017, accepted October 28, 2017, date of publication November 3, 2017,
date of current version December 5, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2769965

Maximum Correntropy Kalman Filter With State Constraints

XI LIU[1], BADONG CHEN[1], (Senior Member, IEEE), HAIQUAN ZHAO[2], (Member, IEEE), JING QIN[3], AND JIUWEN CAO[4], (Member, IEEE)

[1]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China
[3]Center of Smart Health, the School of Nursing, The Hong Kong Polytechnic University, Hong Kong
[4]Institute of Information and Control, Hangzhou Dianzi University, Hangzhou 310018, China

Corresponding author: Badong Chen (chenbd@mail.xjtu.edu.cn)

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351703 and in part by the National Natural Science Foundation of China under Grant 91648208 and Grant 61372152.

ABSTRACT For linear systems, the original Kalman filter under the minimum mean square error (MMSE) criterion is an optimal filter under a Gaussian assumption. However, when the signals follow non-Gaussian distributions, the performance of this filter deteriorates significantly. An efficient way to solve this problem is to use the maximum correntropy criterion (MCC) instead of the MMSE criterion to develop the filters. In a recent work, the maximum correntropy Kalman filter (MCKF) was derived. The MCKF performs very well in filtering heavy-tailed non-Gaussian noise, and its performance can be further improved when some prior information about the system is available (e.g., the system states satisfy some equality constraints). In this paper, to address the problem of state estimation under equality constraints, we develop a new filter, called the MCKF with state constraints, which combines the advantages of the MCC and constrained estimation technology. The performance of the new algorithm is confirmed with two illustrative examples.

INDEX TERMS Kalman filter, robust estimation, maximum correntropy criterion (MCC), state constraints.

I. INTRODUCTION

The well-known Kalman filter (KF) [1]–[3] is a minimum mean square error (MMSE) criterion-based optimal filter for linear systems under Gaussian noise. It has been widely used in many practical applications, including target tracking [4], dynamic optimization [5], signal denoising [6], wire loss monitoring [7] and many others. To date many variants of the Kalman filter have also been proposed to solve the problem of nonlinear estimation. Typical examples include the extended Kalman filter (EKF) [8], the second-order extended Kalman filter (SEKF) [9], the unscented Kalman filter (UKF) [10], and the cubature Kalman filter (CKF) [11].

However, the MMSE criterion-based Kalman filters may not perform well when the systems are subject to non-Gaussian noise [12], which frequently appears in real-world applications. To address the problem of estimation in non-Gaussian noise, some optimization criteria beyond the MMSE have been proposed in the literature. In particular, the maximum correntropy criterion (MCC) has been successfully used in adaptive estimation and filtering under heavy-tailed noise environments [13]–[22]. To improve the performance of the Kalman filter in heavy-tailed non-Gaussian noise environments, we developed the maximum correntropy Kalman filter (MCKF) in [23] by using the MCC instead of the MMSE as the optimization criterion. The MCKF can achieve much better performance than the original KF in many situations.

The performance of the MCKF can be further improved when some prior information about the system is available. For example, the system states sometimes may satisfy certain linear or nonlinear equality constraints. Such systems often occur in engineering applications, including turbofan engine health estimation [24], stereo vision systems [25], hand gesture estimation [26], compartmental models [27] and so on. Considering such systems, some methods have been proposed to solve the problem of state estimation by incorporating the constraints into the estimator. These techniques include model reduction, perfect measurements, estimate projection, probability density function truncation, linear approximation, second-order approximation and so on. For linear constraints, the model reduction approach [28] reduces the equality-constrained filtering to the equivalent unconstrained filtering. The perfect measurements approach [25], [29], [30] augments the measurement equation by

treating the state equality constraints as perfect measurements with zero measurement noise. The estimate projection approach [31] projects the unconstrained estimate onto the constrained surface. The probability density function (PDF) truncation approach [24], [26] truncates the PDF of the state estimate at the constraint edges, and the constrained state estimate is equal to the mean of the truncated PDF. For the nonlinear constraints, the linear approximation [31] method expands the nonlinear state constraints around the current constrained state estimate and keeps only the first-order terms. The second-order approximation method [32] keeps not only the first-order terms but also the second-order terms.

In the present paper, to address the constrained state estimation in non-Gaussian noise environments, a new filtering algorithm called the maximum correntropy Kalman filter with state constraints (MCKF-SC) is proposed. By combining the MCC and constrained estimation technology, which projects the unconstrained MCKF solution onto the state constrained surface, this algorithm can further improve the filtering performance.

The rest of this paper is organized as follows: In Section II, the correntropy and maximum correntropy Kalman filter are briefly introduced. In Section III, four approaches are used to derive the maximum correntropy Kalman filter with state constraints. In Section IV, illustrative examples are presented to show the desirable performance of the new algorithm. Finally, conclusions are given in Section V.

II. PRELIMINARIES

A. CORRENTROPY

Correntropy is a novel concept that can measure the local similarity between two random variables. Given two random variables X and Y, the correntropy is defined by [33]–[35]

$$V(X,Y) = \mathrm{E}\left[\kappa(X,Y)\right] = \int \kappa(x,y)d\mathrm{F}_{XY}(x,y) \tag{1}$$

where E is the expectation operator, $\kappa(\cdot,\cdot)$ denotes a shift-invariant Mercer kernel and $\mathrm{F}_{XY}(x,y)$ is the joint distribution function of (X,Y). In this paper, unless indicated otherwise, the kernel function is the Gaussian kernel:

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

where $e = x - y$, and $\sigma > 0$ denotes the kernel bandwidth. With a Gaussian kernel, the correntropy is insensitive to large errors and hence can be used as a robust cost function in estimation related problems. This estimation criterion is called the maximum correntropy criterion (MCC) [13]–[16], [18]–[22], [34].

B. MAXIMUM CORRENTROPY KALMAN FILTER

The maximum correntropy Kalman filter (MCKF) is a new Kalman-type filter for linear systems that takes advantage of the robustness of correntropy to address impulsive noise.

Consider a linear system with the following state and measurement equations:

$$\mathbf{x}(k) = \mathbf{F}(k-1)\mathbf{x}(k-1) + \mathbf{q}(k-1), \tag{3}$$
$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{r}(k). \tag{4}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ and $\mathbf{y}(k) \in \mathbb{R}^m$ denote the n-dimensional state vector and m-dimensional measurement vector at time k, respectively. $\mathbf{F}(\cdot)$ is the known $n \times n$ state transition matrix and $\mathbf{H}(\cdot)$ is the known $m \times n$ measurement matrix. $\mathbf{q}(k-1)$ and $\mathbf{r}(k)$ are, respectively, the uncorrelated process noise and measurement noise, with zero mean and covariance matrices $\mathbf{Q}(k-1)$ and $\mathbf{R}(k)$.

Like the original KF, the MCKF also includes the following two steps [23]:

1) PREDICT

The prior mean and covariance matrix are computed by

$$\widehat{\mathbf{x}}(k|k-1) = \mathbf{F}(k-1)\widehat{\mathbf{x}}(k-1|k-1), \tag{5}$$
$$\mathbf{P}(k|k-1) = \mathbf{F}(k-1)\mathbf{P}(k-1|k-1)\mathbf{F}^T(k-1) + \mathbf{Q}(k-1). \tag{6}$$

2) UPDATE

Choose a proper kernel bandwidth σ and a small positive threshold ε, and set

$$\begin{aligned}&\begin{bmatrix}\mathbf{P}(k|k-1) & 0\\ 0 & \mathbf{R}(k)\end{bmatrix}\\ &\quad= \begin{bmatrix}\mathbf{B}_p(k|k-1)\mathbf{B}_p^T(k|k-1) & 0\\ 0 & \mathbf{B}_r(k)\mathbf{B}_r^T(k)\end{bmatrix}\\ &\quad= \mathbf{B}(k)\mathbf{B}^T(k)\end{aligned} \tag{7}$$

and construct the following equation

$$\mathbf{D}(k) = \mathbf{W}(k)\mathbf{x}(k) + \mathbf{e}(k) \tag{8}$$

where $\mathbf{D}(k) = \mathbf{B}^{-1}(k)\begin{bmatrix}\widehat{\mathbf{x}}(k|k-1)\\ \mathbf{y}(k)\end{bmatrix}$, $\mathbf{W}(k) = \mathbf{B}^{-1}(k)\begin{bmatrix}\mathbf{I}\\ \mathbf{H}(k)\end{bmatrix}$. The posterior estimation is updated by the fixed-point iteration for $\widehat{\mathbf{x}}(k|k)$:

$$\begin{aligned}&\widehat{\mathbf{x}}^{(t)}(k|k)\\ &\quad= \widehat{\mathbf{x}}(k|k-1) + \widetilde{\mathbf{K}}^{(t-1)}(k)(\mathbf{y}(k) - \mathbf{H}(k)\widehat{\mathbf{x}}(k|k-1))\\ &\text{starting with } \widehat{\mathbf{x}}^{(0)}(k|k)\\ &\quad= \widehat{\mathbf{x}}(k|k-1), \quad t = 1\end{aligned} \tag{9}$$

where

$$\begin{aligned}\widetilde{\mathbf{K}}^{(t-1)}(k) = \widetilde{\mathbf{P}}^{(t-1)}(k|k-1)\mathbf{H}^T(k)\\ \times\left(\mathbf{H}(k)\widetilde{\mathbf{P}}^{(t-1)}(k|k-1)\mathbf{H}^T(k)\right.\\ \left.+\widetilde{\mathbf{R}}^{(t-1)}(k)\right)^{-1}\end{aligned} \tag{10}$$

$$\begin{aligned}\widetilde{\mathbf{P}}^{(t-1)}(k|k-1) = \mathbf{B}_p(k|k-1)\left(\widetilde{\mathbf{C}}_x^{(t-1)}(k)\right)^{-1}\\ \times\mathbf{B}_p^T(k|k-1)\end{aligned} \tag{11}$$

$$\widetilde{\mathbf{R}}^{(t-1)}(k) = \mathbf{B}_r(k)\left(\widetilde{\mathbf{C}}_y^{(t-1)}(k)\right)^{-1}\mathbf{B}_r^T(k) \tag{12}$$

$$\widetilde{\mathbf{C}}_x^{(t-1)}(k) = diag\left(\mathrm{G}_\sigma\left(\widetilde{e}_1^{(t-1)}(k)\right), ..., \mathrm{G}_\sigma\left(\widetilde{e}_n^{(t-1)}(k)\right)\right) \tag{13}$$

$$\widetilde{\mathbf{C}}_y^{(t-1)}(k) = diag\left(\mathrm{G}_\sigma\left(\widetilde{e}_{n+1}^{(t-1)}(k)\right), ..., \mathrm{G}_\sigma\left(\widetilde{e}_{n+m}^{(t-1)}(k)\right)\right) \tag{14}$$

$$\widetilde{e}_i^{(t-1)}(k) = d_i(k) - \mathbf{w}_i(k)\widehat{\mathbf{x}}^{(t-1)}(k|k) \tag{15}$$

in which $\mathbf{B}_p(k|k-1)$ and $\mathbf{B}_r(k)$ are obtained by Cholesky decomposition of $\mathbf{P}(k|k-1)$ and $\mathbf{R}(k)$, respectively, superscript (t) denotes the corresponding value, vector or matrix at the t-th iteration, $d_i(k)$ is the i-th element of $\mathbf{D}(k)$, and $\mathbf{w}_i(k)$ is the i-th row of $\mathbf{W}(k)$. The iteration stops when $\frac{\|\widehat{\mathbf{x}}^{(t)}(k|k)-\widehat{\mathbf{x}}^{(t-1)}(k|k)\|}{\|\widehat{\mathbf{x}}^{(t-1)}(k|k)\|} \leq \varepsilon$, with ε being a small positive value, or iterative number reaches a preset value. Moreover, the posterior covariance is updated by

$$\mathbf{P}(k|k) = \left(\mathbf{I} - \widetilde{\mathbf{K}}(k)\mathbf{H}(k)\right)\mathbf{P}(k|k-1)\left(\mathbf{I} - \widetilde{\mathbf{K}}(k)\mathbf{H}(k)\right)^T + \widetilde{\mathbf{K}}(k)\mathbf{R}(k)\widetilde{\mathbf{K}}^T(k) \tag{16}$$

It should be noted that if the measurements are contaminated by some extremely large noises, the MCKF may face numerical problems since the matrix $\widetilde{\mathbf{C}}_y(k)$ will be nearly singular. Below, we give an approach to solve this problem. Let

$$\eta(k) = \mathbf{y}(k) - \mathbf{H}(k)\widehat{\mathbf{x}}(k|k-1) \tag{17}$$

$$\mathbf{A}(k) = \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k) \tag{18}$$

$$v(k) = \eta^T(k)\mathbf{A}^{-1}(k)\eta(k) \tag{19}$$

If $|v(k)| > \delta$, where $\delta > 0$ is a preset threshold, we only carry out the predict step, that is, $\widehat{\mathbf{x}}(k|k) = \widehat{\mathbf{x}}(k|k-1)$, $\mathbf{P}(k|k) = \mathbf{P}(k|k-1)$. If $|v(k)| \leq \delta$, both the two steps will be carried out.

III. MAXIMUM CORRENTROPY KALMAN FILTER WITH STATE CONSTRAINTS

The MCKF can achieve better performance than the conventional MMSE-based Kalman filters, especially when the underlying dynamic system is disturbed by certain heavy-tailed (or impulsive) non-Gaussian noise. If some prior information about the system is available, the performance of MCKF can be further improved. In the following, we will incorporate the linear or nonlinear constraints on the system states into the MCKF algorithm. For simplicity, we omit the time step k.

A. LINEAR CONSTRAINT

Assume that the state elements are subject to an additional linear constraint

$$\mathbf{M}\mathbf{x} = \mathbf{m} \tag{20}$$

where $\mathbf{M}$ is an $s \times n$ matrix, $\mathbf{m}$ is an $s \times 1$ vector, and $s \leq n$. Moreover, we assume that $\mathbf{M}$ is of full rank.

Next, two methods are proposed to incorporate this linear constraint.

1) ESTIMATE PROJECTION

Estimate projection is a common method to address the problem of constrained filtering. It projects the unconstrained estimate $\widehat{\mathbf{x}}$ onto the constrained surface. The constrained estimate $\overline{\mathbf{x}}$ can be given by the following equation:

$$\overline{\mathbf{x}} = \underset{\mathbf{x}}{argmin}\,(\mathbf{x} - \widehat{\mathbf{x}})^T\mathbf{W}(\mathbf{x} - \widehat{\mathbf{x}}) \tag{21}$$

satisfying

$$\mathbf{M}\mathbf{x} = \mathbf{m} \tag{22}$$

where $\mathbf{W}$ is a positive definite weighting matrix. Then, the Lagrangian function is expressed as

$$L = (\mathbf{x} - \widehat{\mathbf{x}})^T\mathbf{W}(\mathbf{x} - \widehat{\mathbf{x}}) + 2\lambda^T(\mathbf{M}\mathbf{x} - \mathbf{m}) \tag{23}$$

The optimal solution can be found by solving

$$\frac{\partial L}{\partial \mathbf{x}} = 0 \tag{24}$$

$$\frac{\partial L}{\partial \lambda} = 0 \tag{25}$$

Thus, we have

$$\overline{\mathbf{x}} = \widehat{\mathbf{x}} - \mathbf{W}^{-1}\mathbf{M}^T\left(\mathbf{M}\mathbf{W}^{-1}\mathbf{M}^T\right)^{-1}(\mathbf{M}\widehat{\mathbf{x}} - \mathbf{m}) \tag{26}$$

Note that we often set $\mathbf{W} = (\mathbf{P}(k|k))^{-1}$ or $\mathbf{W} = \mathbf{I}$ to obtain the constrained estimate. Table 1 describes all the steps of the MCKF with estimate projection algorithm. The detailed descriptions of the other three methods will be omitted due to the similarity.

2) PROBABILITY DENSITY FUNCTION TRUNCATION

The probability density function (PDF) truncation method truncates the PDF of the state estimate at the constraint edges. The constrained state estimate is equal to the mean of the truncated PDF. The steps of this method are summarized as follows:

First, we define $\overline{\mathbf{x}}^{(t)}$ as the state estimate after the first t constraints of (20) have been enforced, and $\overline{\mathbf{P}}^{(t)}$ as its covariance matrix.

Initialize

$$t = 0, \overline{\mathbf{x}}^{(t)} = \widehat{\mathbf{x}}, \overline{\mathbf{P}}^{(t)} = \mathbf{P}(k|k) \tag{27}$$

Make the following transformation:

$$\mathbf{z}^{(t)} = \rho\mathbf{S}^{-1/2}\mathbf{U}^T(\mathbf{x} - \overline{\mathbf{x}}^{(t)}) \tag{28}$$

where ρ is an orthogonal matrix, $\mathbf{U}$ is an orthogonal matrix, and $\mathbf{S}$ is a diagonal matrix.

Solve the singular value decomposition (SVD) of $\overline{\mathbf{P}}^{(t)}$, that is,

$$\overline{\mathbf{P}}^{(t)} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{29}$$

and obtain the orthogonal matrix ρ by using the Gram-Schmidt orthogonalization that satisfies

$$\rho\mathbf{S}^{1/2}\mathbf{U}^T\mathbf{M}_{t+1}^T = \left[\left(\mathbf{M}_{t+1}\overline{\mathbf{P}}^{(t)}\mathbf{M}_{t+1}^T\right)^{1/2}\ 0\ \cdots\ 0\right]^T \tag{30}$$

where $\mathbf{M}_{t+1}$ denotes the $t+1$-th row of $\mathbf{M}$.

TABLE 1. MCKF with estimate projection algorithm.

1. Initialization

$k=1,\ \widehat{\mathbf{x}}(0|0),\ \mathbf{P}(0|0)$

2. Predict

$\widehat{\mathbf{x}}(k|k-1)=\mathbf{F}(k-1)\widehat{\mathbf{x}}(k-1|k-1)$

$\mathbf{P}(k|k-1)=\mathbf{F}(k-1)\mathbf{P}(k-1|k-1)\mathbf{F}^T(k-1)+\mathbf{Q}(k-1)$

3. Update

$\eta(k)=\mathbf{y}(k)-\mathbf{H}(k)\widehat{\mathbf{x}}(k|k-1)$

$\mathbf{A}(k)=\mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k)+\mathbf{R}(k)$

$v(k)=\eta^T(k)\mathbf{A}^{-1}(k)\eta(k)$

If $|v(k)|>\delta$, then

$\widehat{\mathbf{x}}(k|k)=\widehat{\mathbf{x}}(k|k-1),\ \mathbf{P}(k|k)=\mathbf{P}(k|k-1)$

else

3.1 Initialization

$\widehat{\mathbf{x}}^{(0)}(k|k)=\widehat{\mathbf{x}}(k|k-1),\ t=1$

3.2 Start

$\widehat{\mathbf{x}}^{(t)}(k|k)=\widehat{\mathbf{x}}(k|k-1)+\widetilde{\mathbf{K}}^{(t-1)}(k)(\mathbf{y}(k)-\mathbf{H}(k)\widehat{\mathbf{x}}(k|k-1))$

$\widetilde{\mathbf{K}}^{(t-1)}(k)=\widetilde{\mathbf{P}}^{(t-1)}(k|k-1)\mathbf{H}^T(k)\left(\mathbf{H}(k)\widetilde{\mathbf{P}}^{(t-1)}(k|k-1)\mathbf{H}^T(k)+\widetilde{\mathbf{R}}^{(t-1)}(k)\right)^{-1}$

$\widetilde{\mathbf{P}}^{(t-1)}(k|k-1)=\mathbf{B}_p(k|k-1)\left(\widetilde{\mathbf{C}}_x^{(t-1)}(k)\right)^{-1}\mathbf{B}_p^T(k|k-1)$

$\widetilde{\mathbf{R}}^{(t-1)}(k)=\mathbf{B}_r(k)\left(\widetilde{\mathbf{C}}_y^{(t-1)}(k)\right)^{-1}\mathbf{B}_r^T(k)$

$\widetilde{\mathbf{C}}_x^{(t-1)}(k)=diag\left(G_\sigma\left(\widetilde{e}_1^{(t-1)}(k)\right),...,G_\sigma\left(\widetilde{e}_n^{(t-1)}(k)\right)\right)$

$\widetilde{\mathbf{C}}_y^{(t-1)}(k)=diag\left(G_\sigma\left(\widetilde{e}_{n+1}^{(t-1)}(k)\right),...,G_\sigma\left(\widetilde{e}_{n+m}^{(t-1)}(k)\right)\right)$

$\widetilde{e}_i^{(t-1)}(k)=d_i(k)-\mathbf{w}_i(k)\widehat{\mathbf{x}}^{(t-1)}(k|k)$

If $\frac{\left\|\widehat{\mathbf{x}}^{(t)}(k|k)-\widehat{\mathbf{x}}^{(t-1)}(k|k)\right\|}{\left\|\widehat{\mathbf{x}}^{(t-1)}(k|k)\right\|}\le\varepsilon$, go to 3.3

else

$t=t+1$, go to 3.2

End

End

3.3 Obtain $\widehat{\mathbf{x}}(k|k)$ and

$\mathbf{P}(k|k)=\left(\mathbf{I}-\widetilde{\mathbf{K}}(k)\mathbf{H}(k)\right)\mathbf{P}(k|k-1)\left(\mathbf{I}-\widetilde{\mathbf{K}}(k)\mathbf{H}(k)\right)^T+\widetilde{\mathbf{K}}(k)\mathbf{R}(k)\widetilde{\mathbf{K}}^T(k)$

4. State constraint

$\mathbf{M}\mathbf{x}(k)=\mathbf{m}$

$\overline{\mathbf{x}}(k)=\widehat{\mathbf{x}}(k|k)-\mathbf{W}^{-1}\mathbf{M}^T\left(\mathbf{M}\mathbf{W}^{-1}\mathbf{M}^T\right)^{-1}\left(\mathbf{M}\widehat{\mathbf{x}}(k|k)-\mathbf{m}\right)$

5. $k=k+1$, go to 2.

According to (20), we have

$$\mathbf{M}_{t+1}\mathbf{U}\mathbf{S}^{1/2}\rho^T\mathbf{z}^{(t)}+\mathbf{M}_{t+1}\overline{\mathbf{x}}^{(t)}=m_{t+1} \tag{31}$$

where m_{t+1} denotes the $t+1$-th element of $\mathbf{m}$.

Dividing both sides of (31) by $(\mathbf{M}_{t+1}\overline{\mathbf{P}}^{(t)}\mathbf{M}_{t+1}^T)^{1/2}$ and rearranging it yields

$$\frac{\mathbf{M}_{t+1}\mathbf{U}\mathbf{S}^{1/2}\rho^T}{\left(\mathbf{M}_{t+1}\overline{\mathbf{P}}^{(t)}\mathbf{M}_{t+1}^T\right)^{1/2}}\mathbf{z}^{(t)}=\frac{m_{t+1}-\mathbf{M}_{t+1}\overline{\mathbf{x}}^{(t)}}{\left(\mathbf{M}_{t+1}\overline{\mathbf{P}}^{(t)}\mathbf{M}_{t+1}^T\right)^{1/2}} \tag{32}$$

As shown by (28)∼(30), $\mathbf{z}^{(t)}$ has a mean of zero and a covariance matrix of identity. We define

$$c_{t+1}=\frac{(m_{t+1}-\mathbf{M}_{t+1}\overline{\mathbf{x}}^{(t)})}{\left(\mathbf{M}_{t+1}\overline{\mathbf{P}}^{(t)}\mathbf{M}_{t+1}^T\right)^{1/2}} \tag{33}$$

and have

$$\begin{bmatrix}1&0&\cdots&0\end{bmatrix}\mathbf{z}^{(t)}=c_{t+1} \tag{34}$$

The mean and variance of $\mathbf{z}^{(t+1)}$ are given by

$$\mu=c_{t+1} \tag{35}$$

$$\xi^2=0 \tag{36}$$

and the mean and variance of the transformed state estimate after the first $t+1$ constraints are enforced can be written as

$$\overline{\mathbf{z}}^{(t+1)}=\begin{bmatrix}\mu&0&\cdots&0\end{bmatrix}^T \tag{37}$$

$$Cov(\overline{\mathbf{z}}^{(t+1)})=diag\begin{pmatrix}\xi^2&1&\cdots&1\end{pmatrix} \tag{38}$$

Making the inverse transformation of (28), the mean and variance of the constrained state estimate that match the $t+1$-th constraint can be computed as

$$\overline{\mathbf{x}}^{(t+1)}=\mathbf{U}\mathbf{S}^{1/2}\rho^T\overline{\mathbf{z}}^{(t+1)}+\overline{\mathbf{x}}^{(t)} \tag{39}$$

$$\overline{\mathbf{P}}^{(t+1)}=\mathbf{U}\mathbf{S}^{1/2}\rho^T Cov(\overline{\mathbf{z}}^{(t+1)})\rho\mathbf{S}^{1/2}\mathbf{U}^T \tag{40}$$

The step stops when $t+1=length(\mathbf{m})$, where $length(\mathbf{m})$ denotes the dimension of $\mathbf{m}$.

B. NONLINEAR CONSTRAINT

Now we assume that the states are subject to the following nonlinear constraint

$$\mathrm{g}(\mathbf{x})=\mathbf{m} \tag{41}$$

where g is a nonlinear function with respect to $\mathbf{x}$ and $\mathbf{m}$ is an $s\times 1$ vector.

Below we present two methods to address the nonlinear constraint (41).

1) LINEAR APPROXIMATION

We expand the nonlinear state constraints around the current constrained state estimate $\overline{\mathbf{x}}$, and get

$$\begin{aligned}\mathrm{g}_i(\mathbf{x})-m_i&=\mathrm{g}_i(\overline{\mathbf{x}})+\mathrm{g}_i'(\overline{\mathbf{x}})^T(\mathbf{x}-\overline{\mathbf{x}})\\&\quad+\frac{1}{2!}(\mathbf{x}-\overline{\mathbf{x}})^T\mathrm{g}_i''(\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})+\cdots-m_i\\&=0\end{aligned} \tag{42}$$

where $\mathrm{g}_i(\cdot)$ is the i-th row of $\mathrm{g}(\cdot)$, m_i is the i-th element of $\mathbf{m}$, and $\mathrm{g}_i'(\cdot)$ and $\mathrm{g}_i''(\cdot)$ represent the first and second partial derivatives of $\mathrm{g}_i(\cdot)$.

Retaining only the first-order terms, regardless of whether there are also higher-order terms, and making some rearrangements, we have

$$\mathrm{g}'(\overline{\mathbf{x}})^T\mathbf{x}\approx\mathbf{m}-\mathrm{g}(\overline{\mathbf{x}})+\mathrm{g}'(\overline{\mathbf{x}})^T\overline{\mathbf{x}} \tag{43}$$

which has similar form to (20) by replacing $\mathbf{M}$ and $\mathbf{m}$ with $\mathrm{g}'(\overline{\mathbf{x}})^T$ and $\mathbf{m}-\mathrm{g}(\overline{\mathbf{x}})+\mathrm{g}'(\overline{\mathbf{x}})^T\overline{\mathbf{x}}$. The problem of nonlinear constraints can thus be solved by a similar process of linear constraints.

2) SECOND-ORDER APPROXIMATION

If the state constraint is second-order, it can be viewed as a second-order approximation of the nonlinearity:

$$\begin{aligned} \mathrm{f}(\mathbf{x}(k)) &= \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}^T \begin{bmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{t}^T & t_0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \\ &= \mathbf{x}^T\mathbf{T}\mathbf{x} + \mathbf{t}^T\mathbf{x} + \mathbf{x}^T\mathbf{t} + t_0 = 0 \end{aligned} \tag{44}$$

Like the idea of the constrained Kalman filter of [31], we can project an unconstrained state estimation onto a nonlinear constrained surface:

$$\bar{\mathbf{x}} = \arg\min_{\mathbf{x}} (\mathbf{z} - \mathbf{S}\mathbf{x})^T (\mathbf{z} - \mathbf{S}\mathbf{x}) \tag{45}$$

such that

$$\mathrm{f}(\mathbf{x}) = 0 \tag{46}$$

Clearly, if we set $\mathbf{W} = \mathbf{S}^T\mathbf{S}$ and $\mathbf{z} = \mathbf{S}\widehat{\mathbf{x}}$, equations (45) and (46) become the same as (21) and (22). The solution of the constrained optimization problem in (45) and (46) can be obtained by using the Lagrangian multiplier technique:

$$\bar{\mathbf{x}} = \mathbf{g}^{-1}\mathbf{v}(\mathbf{I} + \beta\Sigma^T\Sigma)^{-1}\mathbf{a}(\beta) \tag{47}$$

$$q(\beta) = \sum_i \frac{a_i^2(\beta)\alpha_i^2}{(1+\beta\alpha_i^2)^2} + 2\sum_i \frac{a_i(\beta)c_j}{1+\beta\alpha_i^2} + t_0 = 0 \tag{48}$$

where $\mathbf{g}$ is an upper right diagonal matrix solved by the Cholesky decomposition of $\mathbf{W} = \mathbf{H}^T\mathbf{H}$:

$$\mathbf{W} = \mathbf{g}^T\mathbf{g} \tag{49}$$

$\mathbf{v}$ is an orthonormal matrix, and Σ is a diagonal matrix with diagonal components denoted by α_i, which can be found by the singular value decomposition (SVD) of the matrix $\mathbf{L}\mathbf{g}^{-1}$ as

$$\mathbf{L}\mathbf{g}^{-1} = \mathbf{U}\Sigma\mathbf{v}^T \tag{50}$$

where $\mathbf{U}$ is the orthonormal matrix of the SVD, $\mathbf{L}$ can be solved by the factorization of $\mathbf{T} = \mathbf{L}^T\mathbf{L}$, and

$$\mathbf{a}(\beta) = [\ldots a_i(\beta)\ldots]^T = \mathbf{v}^T(\mathbf{g}^T)^{-1}(\mathbf{H}^T\mathbf{z} - \beta\mathbf{t}) \tag{51}$$

$$\mathbf{c} = [\ldots c_i\ldots]^T = \mathbf{v}^T(\mathbf{g}^T)^{-1}\mathbf{t} \tag{52}$$

It is difficult to obtain a closed-form solution for the nonlinear equation $q(\beta) = 0$ in (48), so one has to solve it using some numerical root-finding algorithms such as Newton's method. Taking the derivative of $q(\beta)$ with respect to β, we have

$$\begin{aligned} \dot{q}(\beta) = 2\sum_i \frac{a_i(\beta)\dot{a}_i(1+\beta\alpha_i^2)\alpha_i^2 - a_i^2(\beta)\alpha_i^4}{(1+\beta\alpha_i^2)^3} \\ + 2\sum_i \frac{\dot{a}_i c_i(1+\beta\alpha_i^2) - a_i(\beta)c_i\alpha_i^2}{(1+\beta\alpha_i^2)^2} \end{aligned} \tag{53}$$

where

$$\dot{\mathbf{a}} = [\ldots\dot{a}_i\ldots]^T = -\mathbf{v}^T(\mathbf{g}^T)^{-1}\mathbf{t} \tag{54}$$

Then, the solution of β with Newton's method can be updated by

$$\begin{aligned} \beta^{(t+1)} &= \beta^{(t)} - \frac{q(\beta^{(t)})}{\dot{q}(\beta^{(t)})} \\ \text{starting with } \beta^{(0)} &= 0, \quad t = 0 \end{aligned} \tag{55}$$

The iteration stops when $\frac{|\beta^{(t+1)} - \beta^{(t)}|}{|\beta^{(t)}|} \le \varepsilon$, a given small value, or the number of iterations reaches a prespecified number. Substituting the solution of β into (47), one can obtain the constrained optimal solution.

IV. ILLUSTRATIVE EXAMPLES

A. EXAMPLE 1

First, we consider a navigation problem for a land vehicle whose velocity is in direction of θ, which measures clockwise from due east. The state and measurement equations are as follows:

$$\begin{aligned} \mathbf{x}(k) = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}(k-1) + \begin{bmatrix} 0 \\ 0 \\ T\sin\theta \\ T\cos\theta \end{bmatrix} \mathbf{u}(k-1) \\ + \mathbf{q}(k-1) \end{aligned} \tag{56}$$

$$\mathbf{y}(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{x}(k) + \mathbf{r}(k) \tag{57}$$

where T is the discrete time interval, $\mathbf{u}(k)$ denotes the acceleration input, and the state $\mathbf{x}(k) = \begin{bmatrix} x_1(k) & x_2(k) & x_3(k) & x_4(k) \end{bmatrix}^T$ contains the north position, the east position, the north velocity and the east velocity. If the vehicle is on the road with an angle θ, we have

$$\tan\theta = x_1(k)/x_2(k) = x_3(k)/x_4(k) \tag{58}$$

The constrained matrix $\mathbf{M}$ used in our experiments is

$$\mathbf{M} = \begin{bmatrix} 1 & -\tan\theta & 0 & 0 \\ 0 & 0 & 1 & -\tan\theta \end{bmatrix} \tag{59}$$

where $\theta = \pi/3$. The true initial state is $\mathbf{x}(0) = \begin{bmatrix} 0 & 0 & 10\tan\theta & 10 \end{bmatrix}^T$, and the estimated initial state is $\widehat{\mathbf{x}}(0|0) = \begin{bmatrix} 0 & 0 & 10\tan\theta & 10 \end{bmatrix}^T$, and the corresponding covariance matrix is assumed to be $\mathbf{P}(0|0) = diag([900, 900, 4, 4])$. In addition, we consider the following distributions of the noises:

$$\begin{aligned} q_1(k-1) &\sim N(0,4) \\ q_2(k-1) &\sim N(0,4) \\ q_3(k-1) &\sim N(0,1) \\ q_4(k-1) &\sim N(0,1) \\ r_1(k) &\sim 0.9N(0,900) + 0.1N(0,90000) \\ r_2(k) &\sim 0.9N(0,900) + 0.1N(0,90000) \end{aligned}$$

The performance is evaluated by using the following benchmarks:

$$\text{RMSE}_p(k) = \frac{1}{M}\sqrt{\sum_{m=1}^{M}(x_1(k)-\widehat{x}_1(k|k))^2+\sum_{m=1}^{M}(x_2(k)-\widehat{x}_2(k|k))^2}, \quad \text{for } k = 1\ldots K \tag{60}$$

$$\text{ARMSE}_p = \frac{1}{K}\sum_{k=1}^{K}\text{RMSE}_p(k) \tag{61}$$

where M is the total number of Monte Carlo runs and K is the total number of time steps in every Monte Carlo run. Clearly, $\text{RMSE}_p(k)$ denotes the root MSE between the true position and the estimated position at time step k.

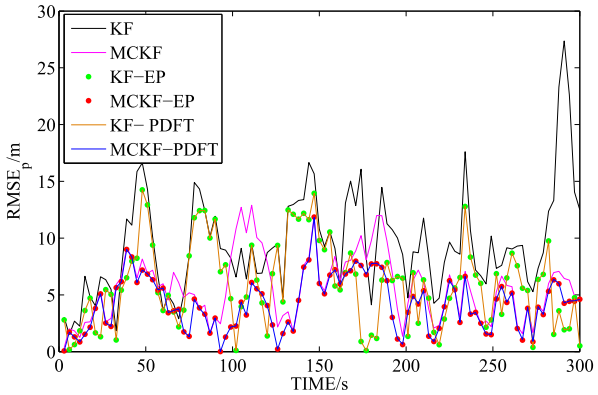


FIGURE 1. RMSE_p for different filters.

TABLE 2. ARMSE_p of different filters.

Filter	ARMSE_p
KF	78.7098
MCKF	55.7389
KF-EP	53.1492
MCKF-EP	39.1217
KF-PDFT	53.1492
MCKF-PDFT	39.1915

In the simulation, the discrete time interval is $T = 3\ s$ and the lasting time is 300 s, and we carry out 100 independent Monte Carlo runs. The kernel bandwidth is set to $\sigma = 2$. Fig. 1 shows the RMSE_p for different filters, including the Kalman filter (KF), the maximum correntropy Kalman filter (MCKF), the Kalman filter with estimate projection (KF-EP), the maximum correntropy Kalman filter with estimate projection (MCKF-EP), the Kalman filter with probability density function truncation (KF-PDFT), and the maximum correntropy Kalman filter with probability density function truncation (MCKF-PDFT). The values of ARMSE_p for different filters are given in Table 2. Since the measurement noises are non-Gaussian, the MCKF family filters are superior to the KF family filters. It can also be seen that the constrained filters can significantly outperform the unconstrained counterparts. The filters that combine the correntropy and constrained estimation technology can achieve excellent performance.

B. EXAMPLE 2

Second, we consider a ground vehicle that travels along a circular road [36]. The turn center is assumed to be the origin of the coordinate system and the turn radius is $r = 100\,m$. The angular turn rate of the vehicle is a constant 5.7 deg/s with an equivalent linear speed of 10 m/s. The true initial state is

$$\begin{aligned}\mathbf{x}(0) &= \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}^T \\ &= \begin{bmatrix} 100\,m & 0\,m/s & 0\,m & 10\,m/s \end{bmatrix}^T\end{aligned}$$

the estimated initial state is assumed to be the same as the true initial state, and corresponding covariance matrix is $\mathbf{P}(0|0) = diag\left(\begin{bmatrix} 5^2 & 1^2 & 5^2 & 1^2 \end{bmatrix}\right)$. The vehicle is tracked by a sensor located at the origin with a sampling interval of $T = 1\ s$. The measurement equation is

$$\mathbf{y}(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\mathbf{x}(k) + \mathbf{r}(k) \tag{62}$$

where the measurement noises are assumed to follow the following distributions

$$r_1(k) \sim 0.8N(0, 9) + 0.2N(0, 900)$$

$$r_2(k) \sim 0.8N(0, 9) + 0.2N(0, 900)$$

Moreover, the state equation is a discrete-time second-order kinematic model, given by

$$\mathbf{x}(k+1) = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} \frac{1}{2}T^2 & 0 \\ T & 0 \\ 0 & \frac{1}{2}T^2 \\ 0 & T \end{bmatrix}\mathbf{q}(k) \tag{63}$$

where the process noises are assumed to be Gaussian, with zero mean and covariance $\mathbf{Q} = diag\left(\begin{bmatrix} 2 & 2 \end{bmatrix}\right)$.

In this example, the performance is evaluated by using the following benchmarks:

$$\text{RMSE}_p(k) = \frac{1}{M}\sqrt{\sum_{m=1}^{M}(x_1(k)-\widehat{x}_1(k|k))^2+\sum_{m=1}^{M}(x_3(k)-\widehat{x}_3(k|k))^2}, \quad \text{for } k = 1\ldots K \tag{64}$$

$$\text{RMSE}_v(k) = \frac{1}{M}\sqrt{\sum_{m=1}^{M}(x_2(k)-\widehat{x}_2(k|k))^2+\sum_{m=1}^{M}(x_4(k)-\widehat{x}_4(k|k))^2}, \quad \text{for } k = 1\ldots K \tag{65}$$

$$\text{ARMSE}_p = \frac{1}{K}\sum_{k=1}^{K}\text{RMSE}_p(k) \tag{66}$$

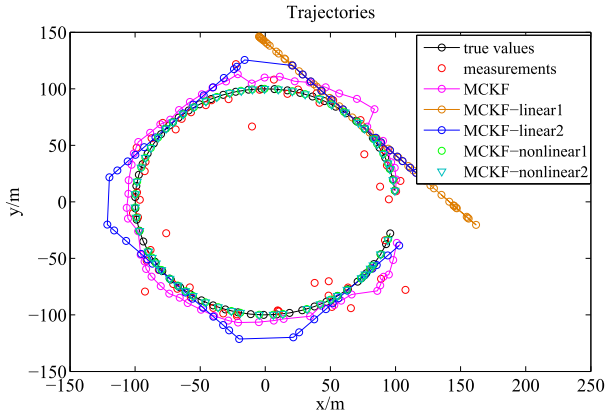


FIGURE 2. Trajectories of different filters.

TABLE 3. ARMSE$_p$ and ARMSE$_v$ of different filters.

Filter	ARMSE$_p$	ARMSE$_v$
KF	10.9234	4.1711
MCKF	10.0285	4.1291
KF-linear1	104.0186	6.9674
KF-linear2	12.5875	4.2057
MCKF-linear1	103.9012	7.0238
MCKF-linear2	11.8897	4.0933
KF-nonlinear1	5.3354	2.6478
KF-nonlinear2	5.3354	1.2226
MCKF-nonlinear1	4.3476	2.4212
MCKF-nonlinear2	4.4060	0.9674

$$\text{ARMSE}_v = \frac{1}{K}\sum_{k=1}^{K}\text{RMSE}_v(k) \tag{67}$$

where M is the total number of Monte Carlo runs and K is the total number of time steps in every Monte Carlo run.

In the simulation, we carry out 100 independent Monte Carlo runs each lasting 60 s. Linear1 denotes the case where a linearizing point at $\theta = 45°$ is selected to cover the entire curved road, and the linearized state constraint at θ is given by

$$\begin{bmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & \cos\theta & 0 & \sin\theta \end{bmatrix}\mathbf{x} = \begin{bmatrix} r \\ 0 \end{bmatrix} \tag{68}$$

Linear2 denotes the case in which four linearized points at $\theta_1 = 45°, \theta_2 = 135°, \theta_3 = 225°, \theta_4 = 315°$ are selected to cover the entire curved road with four linear segments. The switching points are $\theta_5 = 90°$, $\theta_6 = 180°$ and $\theta_7 = 270°$, respectively. Nonlinear1 denotes a nonlinear constraint with respect to the position estimate, given by

$$\text{f}_1(\mathbf{x}(k)) = x_1^2(k) + x_3^2(k) = r^2 \tag{69}$$

Nonlinear2 denotes the constraint with respect to both the position and velocity, and the constrained velocity estimate is obtained by the following projection

$$\bar{\mathbf{v}} = (\hat{\mathbf{v}}\tau)\tau \tag{70}$$

where $\bar{\mathbf{v}} = \begin{bmatrix}\bar{x}_2 & \bar{x}_4\end{bmatrix}^T$ is the constrained velocity estimate, $\hat{\mathbf{v}} = \begin{bmatrix}\hat{x}_2 & \hat{x}_4\end{bmatrix}$ is the unconstrained velocity estimate, and $\tau = \begin{bmatrix}-\sin\theta & \cos\theta\end{bmatrix}^T$ is the constrained unit direction vector associated with the constrained position estimate $\bar{\mathbf{p}} = \begin{bmatrix}\bar{x}_1 & \bar{x}_3\end{bmatrix}^T$ at $\theta = \tan^{-1}(\bar{x}_3/\bar{x}_1)$.

Fig. 2 shows the trajectories of different MCKF family filters, and Table 3 summarizes the ARMSE$_p$ and ARMSE$_v$ of different filters. In the simulation, the kernel bandwidth is set to $\sigma = 2.0$. Again, the MCKF family filters achieve better performance than the KF family filters, and the constrained filters outperform the unconstrained ones. In addition, the second-order nonlinear-approximation-based filters are superior to the linear-approximation-based filters.

V. CONCLUSION

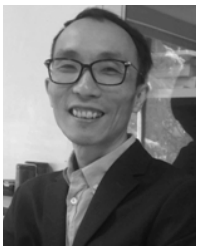
In this paper, in order to address the problem of constrained state estimation in non-Gaussian noise environments, we propose a new filtering algorithm, called MCKF-SC, by combining the MCC and constrained estimation technology. Both linear and nonlinear state constraints are taken into account. Simulation results confirm the superior performance of the proposed algorithm.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.
[2] N. E. Nahi, *Estimation Theory and Applications*. New York, NY, USA: Wiley, 1969.
[3] A. E. Bryson and Y. Ho, *Applied Optimal Control: Optimization, Estimation and Control*. Boca Raton, FL, USA: CRC Press, 1975.
[4] Y. Yang, X. Fan, Z. Zhuo, S. Wang, J. Nan, and Y. Xu, "Amended Kalman filter for maneuvering target tracking," *Chin. J. Electron.*, vol. 25, no. 6, pp. 1166–1171, 2016.
[5] A. Muruganantham, K. C. Tan, and P. Vadakkepat, "Evolutionary dynamic multiobjective optimization via Kalman filter prediction," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2862–2873, Dec. 2016.
[6] M. Narasimhappa, S. L. Sabat, and J. Nayak, "Fiber-optic gyroscope signal denoising using an adaptive robust Kalman filter," *IEEE Sensors J.*, vol. 16, no. 10, pp. 3711–3718, May 2016.
[7] Z. Long, Y. Zheng, C. Li, and Y. He, "Wire loss monitoring in ultrasonic wedge bonding using the Kalman filter algorithm," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 6, no. 1, pp. 153–160, Jan. 2016.
[8] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. New York, NY, USA: Prentice-Hall, 1979.
[9] M. Roth and F. Gustafsson, "An efficient implementation of the second order extended Kalman filter," in *Proc. 14th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2011, pp. 1–6.
[10] 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.
[11] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.
[12] I. C. Schick and S. K. Mitter, "Robust recursive estimation in the presence of heavy-tailed observation noise," *Ann. Stat.*, vol. 22, no. 2, pp. 1045–1080, 1994.
[13] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 880–884, Jul. 2014.
[14] 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.
[15] B. Chen and J. C. Principe, "Maximum correntropy estimation is a smoothed MAP estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 491–494, Aug. 2012.

[16] X. Chen, J. Yang, J. Liang, and Q. Ye, "Recursive robust least squares support vector regression based on maximum correntropy criterion," *Neurocomputing*, vol. 97, pp. 63–73, Nov. 2012.
[17] 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.
[18] 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.
[19] 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.
[20] L. Shi and Y. Lin, "Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1385–1388, Nov. 2014.
[21] B. Chen, Y. Zhu, J. Hu, and J. C. Principe, *System Parameter Identification: Information Criteria and Algorithms*. London, U.K.: Elsevier, 2013.
[22] W. Ma, B. Chen, J. Duan, and H. Zhao, "Diffusion maximum correntropy criterion algorithms for robust distributed estimation," *Digit. Signal Process.*, vol. 58, pp. 10–19, Nov. 2016.
[23] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.
[24] D. Simon and D. L. Simon, "Constrained Kalman filtering via density function truncation for turbofan engine health estimation," *Int. J. Syst. Sci.*, vol. 41, no. 2, pp. 159–171, 2010.
[25] J. Porrill, "Optimal combination and constraints for geometrical sensor data," *Int. J. Robot. Res.*, vol. 7, no. 6, pp. 66–77, 1988.
[26] N. Shimada, Y. Shirai, Y. Kuno, and J. Miura, "Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 268–273.
[27] B. Teixeira, J. Chandrasekar, L. A. B. Tôrres, L. A. Aguirre, and D. S. Bernstein, "State estimation for linear and non-linear equality-constrained systems," *Int. J. Control*, vol. 82, no. 5, pp. 918–936, 2009.
[28] W. Wen and H. F. Durrant-Whyte, "Model-based multi-sensor data fusion," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1992, pp. 1720–1726.
[29] A. T. Alouani and W. D. Blair, "Use of a kinematic constraint in tracking constant speed, maneuvering targets," in *Proc. 30th IEEE Conf. Decision Control*, Dec. 1991, pp. 2055–2058.
[30] L. S. Wang, Y. T. Chiang, and F. R. Chang, "Filtering method for nonlinear systems with constraints," *IEE Proc.-Control Theory Appl.*, vol. 149, no. 6, pp. 525–531, Nov. 2002.
[31] D. Simon and T. L. Chia, "Kalman filtering with state equality constraints," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 128–136, Jan. 2002.
[32] C. Yang and E. Blasch, "Kalman filtering with nonlinear state constraints," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 1, pp. 70–84, Jan. 2009.
[33] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[34] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York, NY, USA: Springer, 2010.
[35] R. J. Bessa, V. Miranda, and J. Gama, "Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1657–1666, Nov. 2009.
[36] X. Feng, Y. Liang, and L. Jiao, "Converted measurement Kalman filter with nonlinear equality constrains," in *Proc. 15th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2012, pp. 1081–1086.

XI LIU is currently pursuing the Ph.D. degree with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include robust estimation and its applications.



BADONG CHEN (M'10–SM'13) received the B.S. and M.S. degrees in control theory and engineering from Chongqing University, in 1997 and 2003, respectively, and the Ph.D. degree in computer science and technology from Tsinghua University in 2008. He was a Post-Doctoral Researcher with Tsinghua University from 2008 to 2010, and a Post-Doctoral Associate with the Computational NeuroEngineering Laboratory, University of Florida, from 2010 to 2012. He visited the Nanyang Technological University as a Visiting Research Scientist in 2015. He is currently a Professor with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University. His research interests are in signal processing, information theory, machine learning, and their applications in cognitive science and engineering. He has authored two books, three chapters, and over 100 papers in various journals and conference proceedings. He is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS and LEARNING SYSTEMS and *The Journal of the Franklin Institute* and has been on the Editorial Board of Entropy.

HAIQUAN ZHAO was born in Suiping, China, in 1974. He received the B.S. degree in applied mathematics and the M.S. and Ph.D. degrees in signal and information processing from the Southwest Jiaotong University, Chengdu, China, in 1998, 2005, and 2011, respectively. Since 2012, he has been a Professor with the School of Electrical Engineering, Southwest Jiaotong University. From 2015 to 2016, he was a Visiting Scholar with the University of Florida, USA. His current research interests include adaptive filtering algorithms, adaptive Volterra filters, nonlinear active noise control, nonlinear system identification, and chaotic signal processing. He has authored or coauthored over 70 journal papers and holds 20 invention patents. He has served as an active reviewer for several IEEE TRANSACTIONS, IET, and other international journals.

JING QIN received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong in 2009. He is currently an Assistant Professor with the School of Nursing, The Hong Kong Polytechnic University. He is also a Key Member of the Centre for Smart Health, SN, PolyU, Hong Kong. His research interests include innovations for healthcare and medicine applications, medical image processing, deep learning, visualization and human-computer interaction, and health informatics.

JIUWEN CAO received the B.Sc. and M.Sc. degrees from the School of Applied Mathematics, University of Electronic Science and Technology of China, Chengdu, China, in 2005 and 2008, respectively, and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore, in 2013. From 2012 to 2013, he was a Research Fellow with NTU. He has been an Associate Professor with Hangzhou Dianzi University, Hangzhou, China, since 2013. His current research interests include machine learning, neural networks, and array signal processing. He has been serving as an Associate Editor for *Multidimensional Systems and Signal Processing* since 2015. Two of his papers have been listed as Top 1% cited papers by Essential Science Indicators in 2016.