Article

Student's *t*-Kernel-Based Maximum Correntropy Kalman Filter

Hongliang Huang [1] and **Hai Zhang** [1,2,*]

1   School of Automation Science and Electrical Engineering, Beihang University, No. 37 Xueyuan Road,
    Haidian District, Beijing 100083, China; huanghl@buaa.edu.cn
2   Science and Technology on Aircraft Control Laboratory, Beihang University, No. 37 Xueyuan Road,
    Haidian District, Beijing 100083, China
*   Correspondence: zhanghai@buaa.edu.cn; Tel.: +86-10-8233-9366

Abstract: The state estimation problem is ubiquitous in many fields, and the common state estimation method is the Kalman filter. However, the Kalman filter is based on the mean square error criterion, which can only capture the second-order statistics of the noise and is sensitive to large outliers. In many areas of engineering, the noise may be non-Gaussian and outliers may arise naturally. Therefore, the performance of the Kalman filter may deteriorate significantly in non-Gaussian noise environments. To improve the accuracy of the state estimation in this case, a novel filter named Student's *t* kernel-based maximum correntropy Kalman filter is proposed in this paper. In addition, considering that the fixed-point iteration method is used to solve the optimal estimated state in the filtering algorithm, the convergence of the algorithm is also analyzed. Finally, comparative simulations are conducted and the results demonstrate that with the proper parameters of the kernel function, the proposed filter outperforms the other conventional filters, such as the Kalman filter, Huber-based filter, and maximum correntropy Kalman filter.

1. Introduction

The state estimation problem is ubiquitous in various applications, such as navigation [1], target tracking [2], and so on [3,4]. The common state estimation method is the Kalman filter (KF), which has been successfully used in many fields [5–7]. For the linear system with additive Gaussian noise, the KF can achieve optimal state estimation. But most systems in real world are usually nonlinear, which limits the application of KF. To solve the state estimation problem of the nonlinear system, many novel filters, such as the extended Kalman filter (EKF) [8], unscented Kalman filter (UKF) [9], quadrature Kalman filter (QKF) [10], cubature Kalman filter (CKF) [11], and so forth have been proposed in the last few decades. The basic idea of EKF is to linearize the nonlinear system by the Taylor expansion technique and truncate the Taylor series at the first-order term. As a result, EKF requires the system model to be differentiable and the Jacobian matrices need to be calculated, resulting in high computational complexity. Besides, for strong nonlinear systems, the first-order linearization will inevitably introduce non-negligible linearization errors, which may lead to the degradation of state estimate performance. The UKF utilizes the unscented transformation (UT) technique to achieve the nonlinear propagation of the mean and covariance of the system state, which avoids the derivation of the Jacobian matrices. Compared with EKF, it has better performance, especially for the strong nonlinear system. To further improve the estimation accuracy, various numerical integration methods are introduced into the filter, such as cubature rule-based filters and quadrature rule-based filters. These methods improve the numerical approximation accuracy of intractable integrals, which leads to more accurate characterization of original probability density functions, correspondingly resulting in an enhanced estimation accuracy [12].

Most of the above filters are based on the mean square error (MSE) criterion and can achieve satisfying state estimation accuracy when the system noises are Gaussian. However, the MSE can only capture the second-order statistics of the noise and is sensitive to large outliers. In many areas of engineering, such as the power system state estimation [13,14], maneuvering target estimation [15] and INS/GNSS integrated navigation [16], the noise may be non-Gaussian and outliers may arise naturally. As a result, the performance of the filters based on MSE may deteriorate significantly. To address these problems, many filters based on non-MSE criteria have been proposed. The Huber-based Kalman filter (HKF) [17] relies on Huber's generalized maximum likelihood methodology which minimizes the combined minimum l_1 and l_2 norm. The residual is bounded by using Huber's function, which suppresses the influence of outliers on state estimation and exhibits strong robustness with respect to deviations from the Gaussian noise. Maximum correntropy Kalman filter (MCKF) [18] is based on the maximum correntropy criterion (MCC) of information-theoretic learning (ITL) [19,20]. Since the correntropy can capture high-order statistics of the noise rather than the common second-order statistics, the state estimation accuracy of MCKF is better than that of KF in the presence of the non-Gaussian noise. To copy with more complicated non-Gaussian noises, minimum error entropy Kalman filter (MEEKF) [21] based on another important learning criterion, minimum error entropy criterion [22,23], was also proposed. The experiment results show the excellent performance of the MEEKF when the underlying system is disturbed by some complicated non-Gaussian noises. Later, various nonlinear filters based on these information-theoretic learning criteria were proposed, such as the maximum correntropy unscented filter [24], minimum error entropy unscented Kalman filter [25], and so forth [26–28].

The performance of the filters based on the information learning criterion heavily depends on the kernel function and its parameters. Currently, most of these filters are based on the Gaussian kernel function. However, in some real applications such as agile targets tracking [29], multipath estimation [30] and measurement outliers from unreliable sensors [31], the systems may be disturbed by heavy-tailed non-Gaussian noises. The Gaussian kernel function may not be the best choice. For the heavy-tailed noises, the Gaussian kernel-based filters may overlook the heavy-tailed properties, which leads to a decrease in the estimation accuracy. Therefore, a novel Cauchy kernel-based maximum correntropy Kalman filter (CKKF) was proposed in the Ref. [32]. Compared with MCKF in which the Gaussian kernel function is used, the performance of CKKF is better for multi-dimensional non-Gaussian noise. Meanwhile, the behavior of the CKKF is consistently stable when a different Cauchy kernel bandwidth is selected.

Recently, a new kernel called Student's *t* kernel function was proposed based on Student's *t* distribution and the Mercer theorem [33]. Compared with the Gaussian kernel function, Student's *t* kernel function can better capture the heavy-tailed features of the noise. The experimental results showed the superiority of the Student's *t* kernel function. Considering the advantages of Student's *t* kernel function, Student's *t* kernel function-based maximum correntropy criteria are applied to the KF algorithm, and a novel Student's *t* kernel-based maximum correntropy Kalman filter algorithm (STKKF) is proposed in this paper. Compared with the Gaussian kernel function, Student's *t* kernel function has two parameters, which can be used to control the shape and the kernel bandwidth of the kernel function, respectively. Thus, it can characterize the distribution of heavy-tailed noise more effectively. The comparative simulations show that with the proper parameters of the kernel function, the accuracy of state estimation of the STKKF outperforms that of conventional algorithms when the noises are heavy-tailed non-Gaussian.

The main contributions of this paper are summarized as follows:

- A novel maximum correntropy Kalman filter is developed in which the Student's *t* kernel function is used to replace the conventional Gaussian kernel function.
- Considering the fixed-point iteration method is used to update the posterior estimates of the state in STKKF, the convergence analysis under a certain condition is given.

- The comparative simulations with other filters are conducted to demonstrate the superiority of STKKF.

The rest of the paper is organized as follows. In Section 2, basic knowledge about the correntropy and Kalman filter are introduced briefly. In Section 3, The STKKF based on Student's *t* kernel maximum correntropy criterion is derived. The convergence of the filter is analyzed in Section 4. To evaluate the performance of STKKF, the comparative simulations are conducted in Section 5. Finally, a discussion is given in Section 6.

2. Preliminaries

2.1. Correntropy

The correntropy was proposed to measure similarity across lags as the autocorrelation of random processes [34], and then was extended to measure the localized similarity of arbitrary two random variables [35]. Let X and Y represent two random variables respectively; then, the correntropy between them can be defined as

$$V(X,Y) = E(\kappa(X,Y)) = \iint \kappa(x,y)p_{X,Y}(x,y)dxdy, \tag{1}$$

where $E(\cdot)$ is the expectation function, $\kappa(X,Y)$ is the kernel function, and $p_{X,Y}(x,y)$ represents the joint probability density function (PDF) of X and Y.

The most widely used Gaussian kernel function is defined as

$$G_\sigma(e) = \kappa(X,Y) = \exp-\left(\frac{\|X-Y\|^2}{2\sigma^2}\right) \qquad (e = X - Y), \tag{2}$$

where σ represents the Gaussian kernel bandwidth.

To better capture the heavy-tailed features in the noise, Student's *t* kernel function [33] is used in this paper to replace the Gaussian kernel function, defined as

$$S_{v,\sigma}(e) = \kappa_{v,\sigma}(X,Y) = \left(1 + \frac{\|X-Y\|^2}{v\sigma^2}\right)^{-\left(\frac{v+2}{2}\right)} \qquad (e = X - Y), \tag{3}$$

where v is used to control the shape of Student's *t* kernel function, and σ is the kernel bandwidth.

In real applications, it is difficult to obtain the joint PDF of random variables. Therefore, the sample mean estimator of correntropy is often used, as shown in the following equation

$$\hat{V}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}\kappa(e_i), \tag{4}$$

where $e_i = x_i - y_i$, with (x_i, y_i), $i = 1, \ldots, N$ being N samples obtained from $p_{X,Y}(x,y)$.

2.2. Kalman Filter

Consider the following linear stochastic system represented by the state-space model

$$\begin{aligned}\mathbf{x}_k &= \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{r}_k,\end{aligned} \tag{5}$$

where k is the discrete time index, $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement vector, $\mathbf{F}_k$ and $\mathbf{H}_k$ are known as state transition matrix and the measurement matrix. $\mathbf{q}_{k-1} \in \mathbb{R}^n$ and $\mathbf{r}_k \in \mathbb{R}^m$ are the mutually independent process and measurement noise, respectively, and satisfy

$$\begin{aligned}E(\mathbf{q}_{k-1}) &= \mathbf{0}, \quad E(\mathbf{r}_k) = \mathbf{0}, \\ E\left(\mathbf{q}_{k-1}\mathbf{q}_{k-1}^T\right) &= \mathbf{Q}_{k-1}, \quad E\left(\mathbf{r}_k\mathbf{r}_k^T\right) = \mathbf{R}_k.\end{aligned} \tag{6}$$

In general, the KF includes the following two steps:

1. One-step state prediction: The priori state estimate $\hat{\mathbf{x}}_{k|k-1}$ and the corresponding error covariance matrix $\mathbf{P}_{k|k-1}$ can be given by

$$
\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \\
\mathbf{P}_{k|k-1} &= \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_{k-1}.
\end{aligned}
\tag{7}
$$

2. Measurement update: The posteriori state estimate $\hat{\mathbf{x}}_{k|k}$ and the corresponding error covariance matrix $\mathbf{P}_{k|k}$ can be given by

$$
\begin{aligned}
\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \right) \\
\mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k{}^T \\
\mathbf{K}_k &= \mathbf{P}_{k|k-1} \mathbf{H}_k{}^T \left(\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k{}^T + \mathbf{R}_k \right)^{-1},
\end{aligned}
\tag{8}
$$

where $\mathbf{K}_k$ is the KF gain matrix.

3. Student's *t* Kernel-Based Maximum Correntropy Kalman Filter

The traditional KF is based on the minimum mean square error (MMSE) criterion and performs well under Gaussian noises. However, when the noises are non-Gaussian or large outliers arise in the measurement, the performance of KF may degrade significantly. Correntropy contains second- and higher order moments of the error and is inherently insensitive to outliers. Therefore, the filters based on the maximum correntropy criterion outperform traditional filters in non-Gaussian noise environments. Meanwhile, these filters are more robust to abnormal measurements. However, the performance of these filters is mainly affected by the kernel function and its parameters. The common Gaussian kernel function may overlook the heavy-tailed properties of heavy-tailed noises, which results in a decrease of the estimation accuracy. To better utilize the heavy-tailed features and improve the estimation accuracy of the system state, Student's *t* kernel function is used to replace the Gaussian kernel function to model and process the heavy-tailed noise.

For the linear system represented by Equation (5), the following equation can be obtained

$$
\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n \times n} \\ \mathbf{H}_k \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k,
\tag{9}
$$

where $\mathbf{v}_k = \left[-\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} \right)^T \quad \mathbf{r}_k^T \right]^T$ and the corresponding covariance matrix can be given as

$$
E\left[\mathbf{v}_k \mathbf{v}_k^T \right] = \begin{bmatrix} \mathbf{P}_{k|k-1} & 0 \\ 0 & \mathbf{R}_k \end{bmatrix} = \begin{bmatrix} \mathbf{B}_p \mathbf{B}_p^T & 0 \\ 0 & \mathbf{B}_r \mathbf{B}_r^T \end{bmatrix},
\tag{10}
$$

where $\mathbf{B}_p$ and $\mathbf{B}_r$ can be obtained from the Cholesky decomposition of $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$, respectively.

When Student's *t* kernel function is employed, the cost function based on the maximum correntropy criterion can be given as

$$
J = \frac{1}{n+m} \left[\sum_{i=1}^{n} S_{v,\sigma}(e_{x,i}) + \sum_{j=1}^{m} S_{v,\sigma}(e_{y,j}) \right],
\tag{11}
$$

where $S_{v,\sigma}(\cdot)$ is Student's t kernel function, and $e_{x,i}$ and $e_{y,j}$ represent the ith and jth element of $\mathbf{e}_x$ and $\mathbf{e}_y$, respectively. The $\mathbf{e}_x$ and $\mathbf{e}_y$ are given by

$$
\begin{aligned}
\mathbf{e}_x &= \mathbf{B}_p^{-1}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right) \\
\mathbf{e}_y &= \mathbf{B}_r^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k).
\end{aligned}
\tag{12}
$$

According to the maximum correntropy criterion, to obtain the optimal state estimation $\hat{\mathbf{x}}_{k|k}$, the following equation should be solved

$$
\frac{\partial J}{\partial \mathbf{x}_k} = \mathbf{0}.
\tag{13}
$$

Substituting Equations (11) and (12) into Equation (13), the following equation can be obtained:

$$
\begin{aligned}
&\sum_{i=1}^{n} \frac{v\sigma^2}{v\sigma^2 + e_{x,i}^2} S_{v,\sigma}(e_{x,i}) \mathbf{B}_{p,i}^{-\mathrm{T}} \mathbf{B}_{p,i}^{-1}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right) \\
&- \sum_{j=1}^{m} \frac{v\sigma^2}{v\sigma^2 + e_{y,j}^2} S_{v,\sigma}(e_{y,j}) \mathbf{H}_k^{\mathrm{T}} \mathbf{B}_{r,j}^{-\mathrm{T}} \mathbf{B}_{r,j}^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) = 0,
\end{aligned}
\tag{14}
$$

where $\mathbf{B}_{p,i}$ and $\mathbf{B}_{r,j}$ represent the ith and the jth row of $\mathbf{B}_p$ and $\mathbf{B}_r$, respectively.

The matrix form of Equation (14) can be expressed as

$$
\mathbf{B}_p^{-\mathrm{T}} \mathbf{\Lambda}_x \mathbf{B}_p^{-1}\left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}\right) - \mathbf{H}_k^{\mathrm{T}} \mathbf{B}_r^{-\mathrm{T}} \mathbf{\Lambda}_y \mathbf{B}_r^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) = 0,
\tag{15}
$$

where

$$
\begin{aligned}
\mathbf{\Lambda}_x &= \mathrm{diag}\left(\frac{v\sigma^2 S_{v,\sigma}(e_{x,1})}{v\sigma^2 + e_{x,1}^2}, \ldots, \frac{v\sigma^2 S_{v,\sigma}(e_{x,n})}{v\sigma^2 + e_{x,n}^2}\right) \\
\mathbf{\Lambda}_y &= \mathrm{diag}\left(\frac{v\sigma^2 S_{v,\sigma}(e_{y,1})}{v\sigma^2 + e_{y,1}^2}, \ldots, \frac{v\sigma^2 S_{v,\sigma}(e_{y,m})}{v\sigma^2 + e_{y,m}^2}\right).
\end{aligned}
\tag{16}
$$

Let

$$
\begin{aligned}
\widetilde{\mathbf{P}}_{k|k-1} &= \mathbf{B}_p \mathbf{\Lambda}_x^{-1} \mathbf{B}_p^{\mathrm{T}} \\
\widetilde{\mathbf{R}}_k &= \mathbf{B}_r \mathbf{\Lambda}_y^{-1} \mathbf{B}_r^{\mathrm{T}}.
\end{aligned}
\tag{17}
$$

Then, Equation (15) can be rewritten as

$$
\left(\widetilde{\mathbf{P}}_{k|k-1}^{-1} + \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{H}_k\right)\mathbf{x}_k = \widetilde{\mathbf{P}}_{k|k-1}^{-1} \hat{\mathbf{x}}_{k|k-1} + \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{y}_k.
\tag{18}
$$

Add and subtract $\mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$ at the right side of Equation (18), where the following equation can be obtained:

$$
\left(\widetilde{\mathbf{P}}_{k|k-1}^{-1} + \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{H}_k\right)\mathbf{x}_k = \left(\widetilde{\mathbf{P}}_{k|k-1}^{-1} + \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{H}_k\right)\hat{\mathbf{x}}_{k|k-1} + \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1}\left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right).
\tag{19}
$$

Then multiplying $\left(\widetilde{\mathbf{P}}_{k|k-1}^{-1} + \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{H}_k\right)^{-1}$ at both sides of Equation (19), we have

$$
\mathbf{x}_k = \hat{\mathbf{x}}_{k|k-1} + \widetilde{\mathbf{K}}_k\left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right),
\tag{20}
$$

where

$$
\widetilde{\mathbf{K}}_k = \left(\widetilde{\mathbf{P}}_{k|k-1}^{-1} + \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{H}_k\right)^{-1} \mathbf{H}_k^{\mathrm{T}} \widetilde{\mathbf{R}}_k^{-1} = \widetilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k \widetilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \widetilde{\mathbf{R}}_k\right)^{-1}.
\tag{21}
$$

Accordingly, the posteriori error covariance matrix $\mathbf{P}_{k|k}$ can be given by

$$\mathbf{P}_{k|k} = \left(\mathbf{I} - \widetilde{\mathbf{K}}_k \mathbf{H}_k\right) \mathbf{P}_{k|k-1} \left(\mathbf{I} - \widetilde{\mathbf{K}}_k \mathbf{H}_k\right)^{\mathrm{T}} + \widetilde{\mathbf{K}}_k \mathbf{R}_k \widetilde{\mathbf{K}}_k^{\mathrm{T}}. \tag{22}$$

It can be seen from Equations (16) and (17) that $\widetilde{\mathbf{K}}_k$ is nonlinear with respect to $\mathbf{x}_k$. Then the fixed-point iterative method is used to solve Equation (20).

In general, the STKKF algorithm can be given as follows:

1. Initialization: The parameters v and σ in the Student's t kernel function are chosen appropriately, and a small number $\varepsilon \in \mathbb{R}^+$ used as an iterative iteration termination condition is given. The initial state $\hat{\mathbf{x}}_0$ and error covariance matrix $\hat{\mathbf{P}}_0$ are set.
2. State prediction: The one-step state prediction $\hat{\mathbf{x}}_{k|k-1}$ and the corresponding error covariance matrix $\mathbf{P}_{k|k-1}$ are the same as those in KF, which can be obtained by Equation (7).
3. Posterior state estimate:
   (a) Calculate the matrix $\mathbf{B}_p$ and $\mathbf{B}_r$ by the Cholesky decomposition of $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$, respectively.
   (b) Let $\hat{\mathbf{x}}_k(l)$ represent the state estimate of the lth fixed-point iteration. At the first iteration, $\hat{\mathbf{x}}_k(l) = \hat{\mathbf{x}}_k(0) = \hat{\mathbf{x}}_{k|k-1}$.
   (c) Calculate the state estimate at the $(l+1)$th iteration by the following equations

$$
\begin{aligned}
\mathbf{e}_x &= \mathbf{B}_p^{-1}\left(\hat{\mathbf{x}}_k(l) - \hat{\mathbf{x}}_{k|k-1}\right) \\
\mathbf{e}_y &= \mathbf{B}_r^{-1}\left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k(l)\right) \\
\boldsymbol{\Lambda}_x &= \mathrm{diag}\left(\frac{v\sigma^2 S_{v,\sigma}(e_{x,1})}{v\sigma^2 + e_{x,1}^2}, \ldots, \frac{v\sigma^2 S_{v,\sigma}(e_{x,n})}{v\sigma^2 + e_{x,n}^2}\right) \\
\boldsymbol{\Lambda}_y &= \mathrm{diag}\left(\frac{v\sigma^2 S_{v,\sigma}(e_{y,1})}{v\sigma^2 + e_{y,1}^2}, \ldots, \frac{v\sigma^2 S_{v,\sigma}(e_{y,m})}{v\sigma^2 + e_{y,m}^2}\right) \\
\widetilde{\mathbf{P}}_{k|k-1} &= \mathbf{B}_p \boldsymbol{\Lambda}_x^{-1} \mathbf{B}_p^{\mathrm{T}} \\
\widetilde{\mathbf{R}}_k &= \mathbf{B}_r \boldsymbol{\Lambda}_y^{-1} \mathbf{B}_r^{\mathrm{T}} \\
\widetilde{\mathbf{K}}_k &= \widetilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} \left(\mathbf{H}_k \widetilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \widetilde{\mathbf{R}}_k\right)^{-1} \\
\hat{\mathbf{x}}_k(l+1) &= \hat{\mathbf{x}}_{k|k-1} + \widetilde{\mathbf{K}}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right).
\end{aligned}
\tag{23}
$$

   (d) Check whether the state estimate in this iteration meets the iteration termination condition by Equation (24). If the termination condition is not met, set $l = l + 1$, return to step (c), and continue the next iteration. Otherwise, set the final state estimate $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_k(l+1)$, and go to step 4.

$$\frac{\|\hat{\mathbf{x}}_k(l+1) - \hat{\mathbf{x}}_k(l)\|}{\|\hat{\mathbf{x}}_k(l)\|} \leq \varepsilon. \tag{24}$$

4. Posterior error covariance update: calculate the corresponding posteriori error covariance matrix $\mathbf{P}_{k|k}$ by Equation (22). Set $k = k + 1$ and return to step 2.

Theorem 1. *If v is fixed, when $\sigma \to \infty$, the STKKF will tend to become the KF.*

Proof. As $\sigma \to \infty$, the matrix $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$ in Equation (16) $\to \mathbf{I}$. Accordingly, $\widetilde{\mathbf{P}}_{k|k-1} \to \mathbf{P}_{k|k-1}$ and $\widetilde{\mathbf{R}}_k \to \mathbf{R}_k$, which means that STKKF reduces to KF. □

Theorem 2. *If σ is fixed, when $v \to \infty$, the STKKF will tend to become the MCKF with bandwidth σ.*

Proof. As $v \to \infty$, the following equation holds:

$$\lim_{v \to \infty} \left(1 + \frac{\|X - Y\|^2}{v\sigma^2} \right)^{-\left(\frac{v+2}{2}\right)} = \lim_{v \to \infty} \left[\left(1 + \frac{\|X - Y\|^2}{v\sigma^2} \right)^{\frac{v\sigma^2}{\|X-Y\|^2}} \right]^{\frac{-(v+2)\|X-Y\|^2}{2v\sigma^2}}$$
$$= \exp\left(-\frac{\|X - Y\|^2}{2\sigma^2} \right), \tag{25}$$

where the equation $\lim_{x \to \infty} (1 + \frac{1}{x}) = e$ is used. Then Student's t kernel function reduces to the Gaussian kernel function, which means that the STKKF tends to become the MCKF. □

4. Convergence Analysis of STKKF

The fixed-point iteration method is used in the STKKF to update the posterior state estimate. To ensure the iterations converge, the convergence of the STKKF is analyzed in this section. The method used is similar to that of the Ref. [36], where only a sufficient condition is given.

The Equation (15) can be rewritten in the following form:

$$\begin{bmatrix} \mathbf{I}^{\mathrm{T}} & \mathbf{H}_k^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{B}_p^{-\mathrm{T}} & \\ & \mathbf{B}_r^{-\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_x & \\ & \boldsymbol{\Lambda}_y \end{bmatrix} \begin{bmatrix} \mathbf{B}_p^{-1} & \\ & \mathbf{B}_r^{-1} \end{bmatrix} \left(\begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix} \mathbf{x}_k - \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix} \right) = \mathbf{0}. \tag{26}$$

Let

$$\mathbf{D} = \begin{bmatrix} \mathbf{B}_p^{-1} & \\ & \mathbf{B}_r^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{B}_p^{-1} & \\ & \mathbf{B}_r^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix},$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Lambda}_x & \\ & \boldsymbol{\Lambda}_y \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} -\mathbf{e}_x \\ \mathbf{e}_y \end{bmatrix} = \mathbf{D} - \mathbf{W}\mathbf{x}_k. \tag{27}$$

Equation (26) can be rewritten in the following form:

$$\mathbf{W}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{W}\mathbf{x}_k - \mathbf{W}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{D} = \mathbf{0}. \tag{28}$$

Then, $\mathbf{x}_k$ can be given as

$$\mathbf{x}_k = \left(\mathbf{W}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{W} \right)^{-1} \mathbf{W}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{D}. \tag{29}$$

Firstly, function $f(\mathbf{x}_k) = \mathbf{x}_k$ is constructed, and by substituting Equation (16) into Equation (29), $f(\mathbf{x}_k)$ can be expressed as

$$f(\mathbf{x}_k) = \left(\sum_{i=1}^{m+n} \frac{v\sigma^2 S_{v,\sigma}(e_i)}{v\sigma^2 + e_i^2} \mathbf{W}_i^T \mathbf{W}_i \right)^{-1} \left(\sum_{i=1}^{m+n} \frac{v\sigma^2 S_{v,\sigma}(e_i)}{v\sigma^2 + e_i^2} D_i \mathbf{W}_i^T \right) = \mathbf{N}_{WW}^{-1} \mathbf{N}_{WD}, \tag{30}$$

where $\mathbf{W}_i$ represents the ith row of matrix $\mathbf{W}$, D_i is the ith component of vector $\mathbf{D}$, and

$$\mathbf{N}_{WW} = \sum_{i=1}^{m+n} \frac{v\sigma^2 S_{v,\sigma}(e_i)}{v\sigma^2 + e_i^2} \mathbf{W}_i^T \mathbf{W}_i, \quad \mathbf{N}_{WD} = \sum_{i=1}^{m+n} \frac{v\sigma^2 S_{v,\sigma}(e_i)}{v\sigma^2 + e_i^2} D_i \mathbf{W}_i^T. \tag{31}$$

The Jacobian matrix of $f(\mathbf{x}_k)$ with respect to $\mathbf{x}_k$, $\nabla_{\mathbf{x}_k} f(\mathbf{x}_k)$, can be expressed as

$$\nabla_{\mathbf{x}_k} f(\mathbf{x}_k) = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}_{k,1}} f(\mathbf{x}_k) & \cdots & \frac{\partial}{\partial \mathbf{x}_{k,n+m}} f(\mathbf{x}_k) \end{bmatrix}, \tag{32}$$

where

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{x}_{k,j}} f(\mathbf{x}_k) &= \frac{\partial}{\partial \mathbf{x}_{k,j}} \mathbf{N}_{WW}^{-1} \mathbf{N}_{WD} \\
&= \mathbf{N}_{WW}^{-1} \left(\frac{\partial}{\partial \mathbf{x}_{k,j}} \mathbf{N}_{WD} \right) \mathbf{N}_{WW}^{-1} \mathbf{N}_{WD} + \mathbf{N}_{WW}^{-1} \left(\frac{\partial}{\partial \mathbf{x}_{k,j}} \mathbf{N}_{WD} \right) \\
&= \mathbf{N}_{WW}^{-1} \left(\frac{\partial}{\partial \mathbf{x}_{k,j}} \mathbf{N}_{WW} \right) f(\mathbf{x}_k) + \mathbf{N}_{WW}^{-1} \left(\frac{\partial}{\partial \mathbf{x}_{k,j}} \mathbf{N}_{WD} \right) \\
&= \mathbf{N}_{WW}^{-1} \left(\sum_{i=1}^{m+n} \frac{v+4}{v\sigma^2} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2} \right)^{-\frac{v+6}{2}} \mathbf{W}_i^T \mathbf{W}_i \right) f(\mathbf{x}_k) \\
&\quad + \mathbf{N}_{WW}^{-1} \left(\sum_{i=1}^{m+n} \frac{v+4}{v\sigma^2} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2} \right)^{-\frac{v+6}{2}} D_i \mathbf{W}_i^T \right).
\end{aligned}
\tag{33}
$$

Then, the following theorem holds.

Theorem 3. *If parameter v is fixed, $\beta > \xi$, and $\sigma > \max(\sigma*, \sigma^+)$, where $\sigma*$ is the solution of $\varphi(v,\sigma) = \beta$, σ^+ is the solution of $\psi(v,\sigma) = \alpha$ $(0 < \alpha < 1)$, then for $\forall \mathbf{x}_k \in \{\mathbb{R}^n : \|\mathbf{x}_k\|_1 < \beta\}$, Equation (34) holds. The expression of ξ, $\varphi(v,\sigma)$, and $\psi(v,\sigma)$ are shown in Equations (35), (36) and (37), respectively.*

$$
\begin{aligned}
\|f(\mathbf{x}_k)\|_1 &\le \beta \qquad (\beta > \xi) \\
\left\| \frac{\partial}{\partial \mathbf{x}_k} f(\mathbf{x}_k) \right\|_1 &\le \alpha \quad (0 < \alpha < 1)
\end{aligned}
\tag{34}
$$

$$
\xi = \frac{\sqrt{n} \sum_{i=1}^{m+n} |D_i| \|\mathbf{W}_i^T\|_1}{\lambda_{\min} \left(\sum_{i=1}^{m+n} \mathbf{W}_i^T \mathbf{W}_i \right)}
\tag{35}
$$

$$
\varphi(v,\sigma) = \frac{\sqrt{n} \sum_{i=1}^{m+n} |D_i| \|\mathbf{W}_i^T\|_1}{\lambda_{\min} \left(\sum_{i=1}^{m+n} \left(1 + \frac{(|D_i| + \beta\|\mathbf{W}_i\|_1)^2}{v\sigma^2} \right)^{-\left(\frac{v+4}{2} \right)} \mathbf{W}_i^T \mathbf{W}_i \right)}
\tag{36}
$$

$$
\psi(v,\sigma) = \frac{(v+4)\sqrt{n} \sum_{i=1}^{m+n} \left[(|D_i| + \beta\|\mathbf{W}_i\|_1) \|\mathbf{W}_i\|_1 \left(\beta\|\mathbf{W}_i^T \mathbf{W}_i\|_1 + \|D_i \mathbf{W}_i^T\|_1 \right) \right]}{v\sigma^2 \lambda_{\min} \left(\sum_{i=1}^{m+n} \left(1 + \frac{(|D_i| + \beta\|\mathbf{W}_i\|_1)^2}{v\sigma^2} \right)^{-\left(\frac{v+4}{2} \right)} \mathbf{W}_i^T \mathbf{W}_i \right)}
\tag{37}
$$

Proof.

$$
\|f(\mathbf{x}_k)\|_1 = \left\| \mathbf{N}_{WW}^{-1} \mathbf{N}_{WD} \right\|_1 \le^{(a)} \left\| \mathbf{N}_{WW}^{-1} \right\|_1 \|\mathbf{N}_{WD}\|_1,
\tag{38}
$$

where $\|\cdot\|_p$ is the l_p-norm of a vector or induced matrix norm defined by $\|\mathbf{A}\|_p = \max_{\|\mathbf{x}\|_p \ne 0} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$, and (a) comes from the compatibility of the matrix norm and vector norm.

According to the matrix theory, the following equation holds:

$$
\left\| \mathbf{N}_{WW}^{-1} \right\|_1 \le \sqrt{n} \left\| \mathbf{N}_{WW}^{-1} \right\|_2 = \sqrt{n} \lambda_{\max} \left(\mathbf{N}_{WW}^{-1} \right),
\tag{39}
$$

where $\lambda_{\max}(\cdot)$ represents the maximum eigenvalue of the matrix.

$$
\begin{aligned}
\lambda_{\max}\left(\mathbf{N}_{WWW}^{-1}\right) \quad &= \frac{1}{\lambda_{\min}(\mathbf{N}_{WW})} \\
&= \frac{1}{\lambda_{\min}\left(\sum_{i=1}^{m+n} \frac{v\sigma^2}{v\sigma^2+e_i^2} S_{v,\sigma}(e_i)\mathbf{W}_i^T\mathbf{W}_i\right)} \\
&= \frac{1}{\lambda_{\min}\left(\sum_{i=1}^{m+n} \left(1+\frac{e_i^2}{v\sigma^2}\right)^{-\left(\frac{v+4}{2}\right)}\mathbf{W}_i^T\mathbf{W}_i\right)} \\
&= \frac{1}{\lambda_{\min}\left(\sum_{i=1}^{m+n} \left(1+\frac{(D_i-\mathbf{W}_i\mathbf{x}_k)^2}{v\sigma^2}\right)^{-\left(\frac{v+4}{2}\right)}\mathbf{W}_i^T\mathbf{W}_i\right)} \\
&\leq^{(b)} \frac{1}{\lambda_{\min}\left(\sum_{i=1}^{m+n} \left(1+\frac{(|D_i|+\beta\|\mathbf{W}_i\|_1)^2}{v\sigma^2}\right)^{-\left(\frac{v+4}{2}\right)}\mathbf{W}_i^T\mathbf{W}_i\right)},
\end{aligned}
\tag{40}
$$

where (b) comes from $\|\mathbf{x}_k\|_1 < \beta$.

Similarly,

$$
\begin{aligned}
\|\mathbf{N}_{WD}\|_1 &= \left\|\sum_{i=1}^{m+n} \frac{v\sigma^2}{v\sigma^2+e_i^2} S_{v,\sigma}(e_i)D_i\mathbf{W}_i^T\right\|_1 \\
&= \left\|\sum_{i=1}^{m+n} \left(1+\frac{e_i^2}{v\sigma^2}\right)^{-\left(\frac{v+4}{2}\right)} D_i\mathbf{W}_i^T\right\|_1 \\
&\leq^{(c)} \sum_{i=1}^{m+n} \left\|D_i\mathbf{W}_i^T\right\|_1 \\
&\leq \sum_{i=1}^{m+n} |D_i| \left\|\mathbf{W}_i^T\right\|_1,
\end{aligned}
\tag{41}
$$

where (c) is because $\left(1+\frac{e_i^2}{v\sigma^2}\right)^{-\left(\frac{v+4}{2}\right)} < 1$ for $\forall e_i$ and the convexity of the l_1 norm.

According to Equations (38), (39) and (41), the following equation holds:

$$
\|f(\mathbf{x}_k)\|_1 \leq \varphi(v,\sigma) = \frac{\sqrt{n}\sum_{i=1}^{m+n} |D_i| \left\|\mathbf{W}_i^T\right\|_1}{\lambda_{\min}\left(\sum_{i=1}^{m+n} \left(1+\frac{(|D_i|+\beta\|\mathbf{W}_i\|_1)^2}{v\sigma^2}\right)^{-\left(\frac{v+4}{2}\right)}\mathbf{W}_i^T\mathbf{W}_i\right)}.
\tag{42}
$$

If the parameter v is fixed, then $\varphi(v,\sigma)$ is the monotonically decreasing function of σ, and then we have

$$
\begin{aligned}
&\lim_{\sigma\to 0^+} \varphi(v,\sigma) = \infty \\
&\lim_{\sigma\to\infty} \varphi(v,\sigma) = \frac{\sqrt{n}\sum_{i=1}^{m+n} |D_i| \left\|\mathbf{W}_i^T\right\|_1}{\lambda_{\min}\left(\sum_{i=1}^{m+n} \mathbf{W}_i^T\mathbf{W}_i\right)} = \xi.
\end{aligned}
\tag{43}
$$

Therefore, for $\forall \beta > \xi$, $\exists$, the unique $\sigma^* \in (0,\infty)$, s.t. $\varphi(v,\sigma^*) = \beta$. When $\sigma > \sigma^*$, $\varphi(v,\sigma) \leq \beta$, that is,

$$
\|f(\mathbf{x}_k)\|_1 \leq \beta.
\tag{44}
$$

According to the matrix theory, to prove $\left\|\nabla_{\mathbf{x}_k} f(\mathbf{x}_k)\right\|_1 \leq \alpha$, we just need to prove $\left\|\frac{\partial}{\partial \mathbf{x}_{k,j}} f(\mathbf{x}_k)\right\|_1 \leq \alpha$ for $\forall j$. The Equation (33) is rewritten here:

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{x}_{k,j}} f(\mathbf{x}_k) = {} & \mathbf{N}_{WW}^{-1}\left(\sum_{i=1}^{m+n} \frac{v+4}{v\sigma^2} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2}\right)^{-\frac{v+6}{2}} \mathbf{W}_i^T \mathbf{W}_i\right) f(\mathbf{x}_k) \\
& + \mathbf{N}_{WW}^{-1}\left(\sum_{i=1}^{m+n} \frac{v+4}{v\sigma^2} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2}\right)^{-\frac{v+6}{2}} D_i \mathbf{W}_i^T\right).
\end{aligned}
\tag{45}
$$

The following equation can be derived:

$$
\begin{aligned}
& \left\| \mathbf{N}_{WW}^{-1}\left(\sum_{i=1}^{m+n} \frac{v+4}{v\sigma^2} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2}\right)^{-\frac{v+6}{2}} \mathbf{W}_i^T \mathbf{W}_i\right) f(\mathbf{x}_k) \right\|_1 \\
& \leq \frac{v+4}{v\sigma^2} \left\| \mathbf{N}_{WW}^{-1}\right\|_1 \left\|\sum_{i=1}^{m+n} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2}\right)^{-\frac{v+6}{2}} \mathbf{W}_i^T \mathbf{W}_i\right\|_1 \left\| f(\mathbf{x}_k)\right\|_1 \\
& \leq^{(d)} \frac{(v+4)\beta}{v\sigma^2} \left\| \mathbf{N}_{WW}^{-1}\right\|_1 \left\|\sum_{i=1}^{m+n} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2}\right)^{-\frac{v+6}{2}} \mathbf{W}_i^T \mathbf{W}_i\right\|_1 \\
& \leq^{(e)} \frac{(v+4)\beta}{v\sigma^2} \left\| \mathbf{N}_{WW}^{-1}\right\|_1 \left[\sum_{i=1}^{m+n} (|D_i| + \beta\|\mathbf{W}_i\|_1)\|\mathbf{W}_i\|_1 \left\|\mathbf{W}_i^T \mathbf{W}_i\right\|_1\right],
\end{aligned}
\tag{46}
$$

where (d) comes from that when $\sigma > \sigma^*$, $\|f(\mathbf{x}_k)\|_1 \leq \beta$, (e) is because of the convexity of the l_1 norm, and $\left|e_i W_i^j\right| = \left|(D_i - \mathbf{W}_i \mathbf{x}_k) W_i^j\right| \leq (|D_i| + \beta\|\mathbf{W}_i\|_1)\|\mathbf{W}_i\|_1$.

Similarly,

$$
\begin{aligned}
& \left\| \mathbf{N}_{WW}^{-1}\left(\sum_{i=1}^{m+n} \frac{v+4}{v\sigma^2} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2}\right)^{-\frac{v+6}{2}} D_i \mathbf{W}_i^T\right) \right\|_1 \\
& \leq \frac{v+4}{v\sigma^2} \left\| \mathbf{N}_{WW}^{-1}\right\|_1 \left\|\sum_{i=1}^{m+n} e_i W_i^j \left(1 + \frac{e_i^2}{v\sigma^2}\right)^{-\frac{v+6}{2}} D_i \mathbf{W}_i^T\right\|_1 \\
& \leq \frac{v+4}{v\sigma^2} \left\| \mathbf{N}_{WW}^{-1}\right\|_1 \left[\sum_{i=1}^{m+n} (|D_i| + \beta\|\mathbf{W}_i\|_1)\|\mathbf{W}_i\|_1 \left\|D_i \mathbf{W}_i^T\right\|_1\right].
\end{aligned}
\tag{47}
$$

According to Equations (40), (46) and (47), the following equation can be obtained:

$$
\begin{aligned}
& \left\|\frac{\partial}{\partial \mathbf{x}_{k,j}} f(\mathbf{x}_k)\right\|_1 \leq \psi(v, \sigma) \\
& = \frac{(v+4)\sqrt{n} \sum_{i=1}^{m+n} \left[(|D_i| + \beta\|\mathbf{W}_i\|_1)\|\mathbf{W}_i\|_1 \left(\beta\|\mathbf{W}_i^T \mathbf{W}_i\|_1 + \|D_i \mathbf{W}_i^T\|_1\right)\right]}{v\sigma^2 \lambda_{\min}\left(\sum_{i=1}^{m+n} \left(1 + \frac{(|D_i| + \beta\|\mathbf{W}_i\|_1)^2}{v\sigma^2}\right)^{-\left(\frac{v+4}{2}\right)} \mathbf{W}_i^T \mathbf{W}_i\right)}.
\end{aligned}
\tag{48}
$$

Additionally, if the parameter v is fixed, then $\psi(v, \sigma)$ is the monotonically decreasing function of σ; then, we have

$$
\begin{aligned}
\lim_{\sigma \to 0^+} \psi(v, \sigma) &= \infty \\
\lim_{\sigma \to \infty} \psi(v, \sigma) &= 0.
\end{aligned}
\tag{49}
$$

Therefore, for $\forall \, \alpha \in (0, 1)$, $\exists$ the unique σ^+, s.t. $\varphi(v, \sigma^+) = \alpha$. When $\sigma > \sigma^+$, the following equation holds:

$$\left\| \frac{\partial}{\partial \mathbf{x}_{k,j}} f(\mathbf{x}_k) \right\|_1 \leq \alpha. \tag{50}$$

Based on the above derivation, we conclude that when the parameter v is fixed, $\sigma > \max(\sigma*, \sigma^+)$, and $\mathbf{x}_k \in \{\mathbb{R}^n : \|\mathbf{x}_k\|_1 < \beta\}$, the following equations hold:

$$\begin{aligned} \|f(\mathbf{x}_k)\|_1 &\leq \beta && (\beta > \xi) \\ \left\| \frac{\partial}{\partial \mathbf{x}_k} f(\mathbf{x}_k) \right\|_1 &\leq \alpha && (0 < \alpha < 1). \end{aligned} \tag{51}$$

The theorem is proved completely. □

By Theorem 3 and the Banach Fixed-Point Theorem [37], if the l_1-norm of the initial iteration point $\|\hat{\mathbf{x}}_k(0)\|_1 \leq \beta$, then STKKF will surely converge to the unique point in range $\mathbf{x}_k \in \{\mathbb{R}^n : \|\mathbf{x}_k\|_1 \leq \beta\}$, provided that the kernel bandwidth σ is larger than a certain value. The implementation pseudocode of the STKKF is shown in Algorithm 1.

Algorithm 1: The implementation pseudocode for one time-step of the STKKF.

Inputs: $\hat{\mathbf{x}}_{k-1|k-1}$, $\mathbf{P}_{k-1|k-1}$, $\mathbf{Q}_{k-1}$, $\mathbf{R}_k$, v, σ, ε.

Time update:

1. $\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}$.

2. $\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_{k-1}$.

Measurement update:

1. $\mathbf{B}_p = Chol(\mathbf{P}_{k|k-1})$, $\mathbf{B}_r = Chol(\mathbf{R}_k)$, $\hat{\mathbf{x}}_k(l = 0) = \hat{\mathbf{x}}_{k|k-1}$.

2. $\mathbf{e}_x = \mathbf{B}_p^{-1}\left(\hat{\mathbf{x}}_k(l) - \hat{\mathbf{x}}_{k|k-1} \right)$, $\mathbf{e}_y = \mathbf{B}_r^{-1}(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k(l))$.

3. $\mathbf{\Lambda}_x = \mathrm{diag}\left(\frac{v\sigma^2 S_{v,\sigma}(e_{x,1})}{v\sigma^2 + e_{x,1}^2}, \dots, \frac{v\sigma^2 S_{v,\sigma}(e_{x,n})}{v\sigma^2 + e_{x,n}^2} \right)$,

   $\mathbf{\Lambda}_y = \mathrm{diag}\left(\frac{v\sigma^2 S_{v,\sigma}(e_{y,1})}{v\sigma^2 + e_{y,1}^2}, \dots, \frac{v\sigma^2 S_{v,\sigma}(e_{y,m})}{v\sigma^2 + e_{y,m}^2} \right)$.

4. $\widetilde{\mathbf{P}}_{k|k-1} = \mathbf{B}_p \mathbf{\Lambda}_x^{-1} \mathbf{B}_p^{\mathrm{T}}$, $\widetilde{\mathbf{R}}_k = \mathbf{B}_r \mathbf{\Lambda}_y^{-1} \mathbf{B}_r^{\mathrm{T}}$.

5. $\widetilde{\mathbf{K}}_k = \widetilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} \left(\mathbf{H}_k \widetilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \widetilde{\mathbf{R}}_k \right)^{-1}$.

6. $\hat{\mathbf{x}}_k(l+1) = \hat{\mathbf{x}}_{k|k-1} + \widetilde{\mathbf{K}}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \right)$.

7. Check $\frac{\|\hat{\mathbf{x}}_k(l+1) - \hat{\mathbf{x}}_k(l)\|}{\|\hat{\mathbf{x}}_k(l)\|} \leq \varepsilon$, where if the termination condition is met, then set $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_k(l+1)$ and go to Step 8; otherwise, set $l = l+1$ and return to Step 2, and continue the next iteration.

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

Outputs: $\hat{\mathbf{x}}_{k|k}$, $\mathbf{P}_{k|k}$.

---

5. Simulations and Results

In this section, simulations are conducted to demonstrate the performance of STKKF. The results of KF, HKF, MCKF, and STKKF are compared when different kinds of noise distribution exist.

The benchmark navigation problem is considered [38]. The dynamics and measurement model are given as follows:

$$\mathbf{x}_k = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{q}_k$$

$$\mathbf{y}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{r}_k,$$

(52)

where Δt is the sample period, and $\mathbf{q}_k$ and $\mathbf{r}_k$ represent the process noise and measurement noise respectively. The first two components of state vector $\mathbf{x}_k \in \mathbb{R}^4$ represent the north and east position of a land vehicle, and the last two components are the corresponding north velocity and east velocity. The position of the vehicle is measured directly by a device.

In the simulation, the sample period Δt is 1 s, and the initial values of the true state $\mathbf{x}_0$, estimated state $\hat{\mathbf{x}}_{0|0}$, and error covariance matrix $\mathbf{P}_{0|0}$ are assumed to be

$$\mathbf{x}_0 = [1, 1, 1, 1]$$

$$\hat{\mathbf{x}}_{0|0} = [1, 1, 1, 1] + N\left(\mathbf{0}, \mathbf{P}_{0|0}\right)$$

$$\mathbf{P}_{0|0} = \text{diag}(0.1, 0.1, 0.1, 0.1).$$

(53)

Two different cases with different kinds of processes and measurement noises are considered in this simulation as follows:

1. The process noise and measurement noise are both Gaussian noises.

$$\mathbf{q}_k \sim N\left(\mathbf{0}, \text{diag}\left(0.1^2, 0.1^2, 0.1^2, 0.1^2\right)\right)$$

$$\mathbf{r}_k \sim N(\mathbf{0}, \text{diag}(1, 1)).$$

(54)

2. The process noise is a Gaussian distribution and the measurement noise is a Gaussian mixture noise.

$$\mathbf{q}_k \sim N\left(\mathbf{0}, \text{diag}\left(0.1^2, 0.1^2, 0.1^2, 0.1^2\right)\right)$$

$$\mathbf{r}_k \sim 0.9N(\mathbf{0}, \text{diag}(1, 1)) + 0.1N\left(\mathbf{0}, \text{diag}\left(10^2, 10^2\right)\right).$$

(55)

The proposed STKKF and the other filters are coded with MATLAB, and the simulations are run on a computer with Intel Core i7-3540M CPU at 3.0 GHz. The time steps in the simulation is 200. For each case, 100 Monte Carlo simulations are implemented to quantify estimation performance. The performance of the filters is evaluated by the root mean square error (RMSE) and average RMSE (ARMSE). The RMSE and ARMSE in position are defined as:

$$\text{RMSE}_{\text{pos}}(i) = \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left(\left(x_{i,1}^j - \hat{x}_{i,1}^j\right)^2 + \left(x_{i,2}^j - \hat{x}_{i,2}^j\right)^2 \right)}$$

$$\text{ARMSE}_{\text{pos}} = \frac{1}{K} \sum_{i=1}^{K} \text{RMSE}_{\text{pos}}(i),$$

(56)

where M is the number of Monte Carlo simulations and K is the time steps in every Monte Carlo simulation. $(x_{i,1}^j, x_{i,2}^j)$ and $(\hat{x}_{i,1}^j, \hat{x}_{i,2}^j)$ are the true position and estimated position at the ith time step in the jth simulation. The RMSE and ARMSE in velocity are similarly defined.

In case 1, the process and measurement noise are both Gaussian noises. Theoretically, KF should be the best estimator. To demonstrate the relationship between the KF and STKKF, these two filters were studied in this case. The ARMSE of the position and velocity

estimates are listed in Table 1. Meanwhile, the average iteration number required for the STKKF to converge and the average implementation times of the filters for one time step are also listed. The iteration termination condition is given by Equation (24). Here, ε is set as 10^{-4}. In addition, the RMSEs of the position and velocity of KF and STKKF with different kernel parameters are also plotted in Figure 1a and Figure 1b, respectively.

Table 1. ARMSEs of position and velocity, average iteration numbers, and average implementation times for one time step of the proposed STKKF and KF when both the process noise and measurement noise are Gaussian noises.

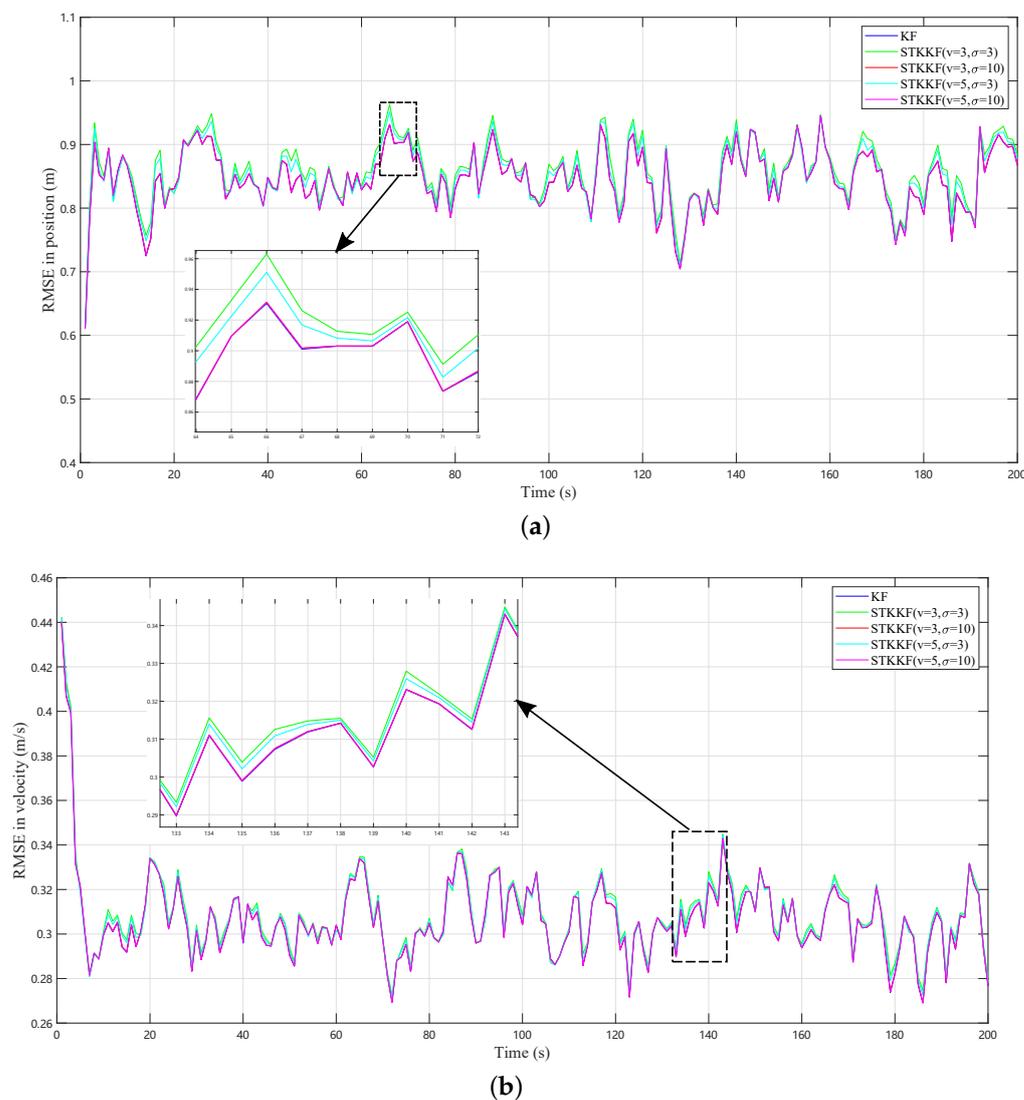| Filters	$ARMSE_{pos}$ (m)	$ARMSE_{vel}$ (m/s)	Average Iteration Number	Time (ms)
KF	0.8459	0.3037	0	0.0140
STKKF($v = 3, \sigma = 3$)	0.8585	0.3091	3.5356	0.1386
STKKF($v = 3, \sigma = 10$)	0.8461	0.3073	2.3422	0.1008
STKKF($v = 3, \sigma = 50$)	0.8459	0.3073	2.0124	0.0887
STKKF($v = 5, \sigma = 3$)	0.8528	0.3083	2.3100	0.1314
STKKF($v = 5, \sigma = 10$)	0.8460	0.3073	2.2833	0.0954
STKKF($v = 5, \sigma = 50$)	0.8459	0.3073	2.0083	0.0877

(**a**)

(**b**)

Figure 1. RMSE results of KF and STKKF in case 1 where RMSE is taken over 100 Monte Carlo simulations: (**a**) position. (**b**) velocity.

In case 2, the process noises are still Gaussian but the measurement noise is a heavy-tailed (impulsive) non-Gaussian noise, with a mixed-Gaussian distribution. The ARMSEs of the position and velocity estimate of different filters, average iteration numbers for MCKF and STKKF, and the average implementation times of the filters for one time step are listed in Table 2. In MCKF, the iteration termination parameter is set to be the same as that of STKKF, that is, 10^{-4}. The RMSEs of the position and velocity of different algorithms are also plotted in Figure 2a and Figure 2b, respectively. It should be noted that for the MCKF and STKKF with different kernel parameters listed in Table 2, only partial models of them are plotted in the figures to maintain the clarity of the plot. It should be pointed out that the noise parameters of KF in the Table 2 are set with the true noise covariance of mixture distribution in Equation (55), and hence, KF is the best linear estimator in the MSE sense. The parameter used in the HKF is set to ensure that the best estimation accuracy is obtained.
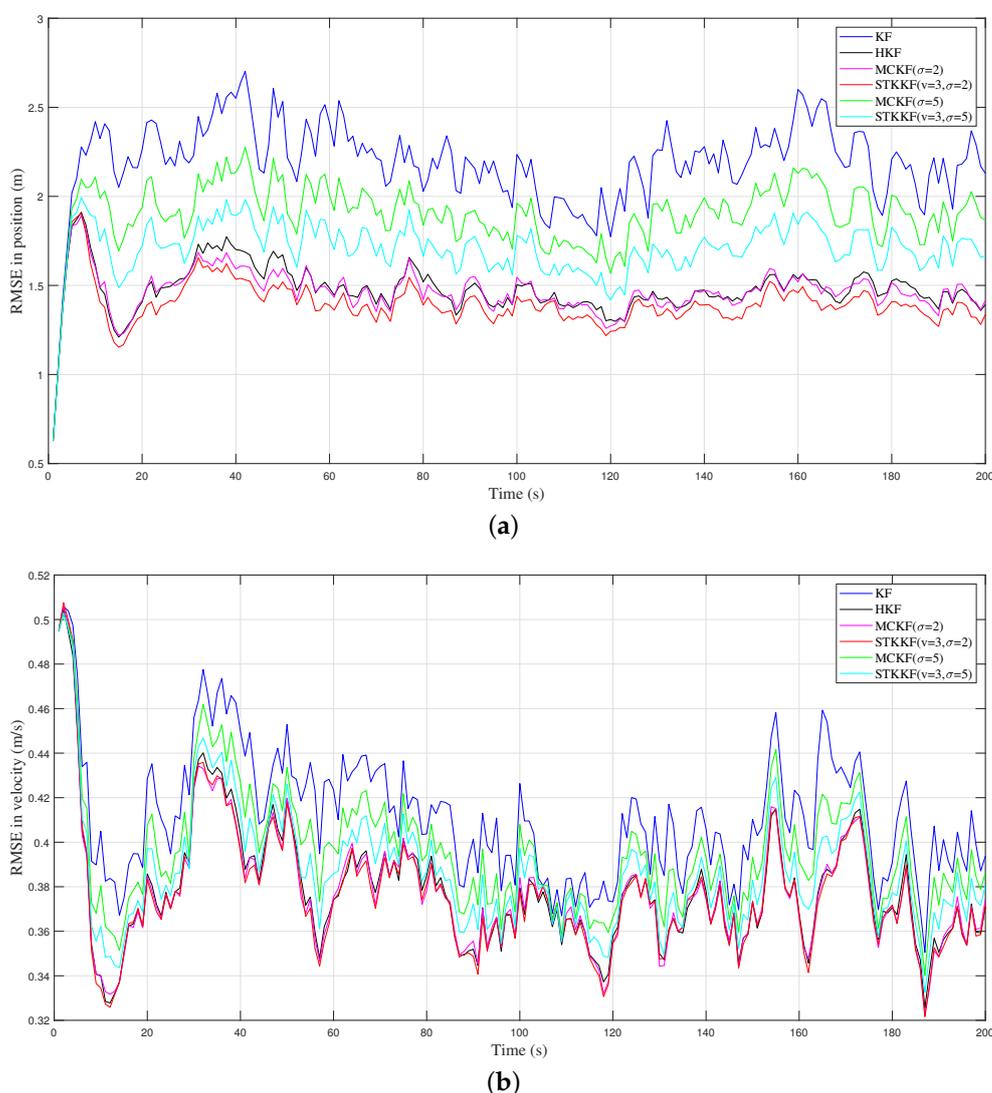

Figure 2. RMSE results of different filters in case 2 where RMSE is taken over 100 Monte Carlo simulations: (**a**) position. (**b**) velocity.

Table 2. ARMSEs of position and velocity, average iteration numbers, and average implementation times for one time step of different algorithms when the measurement noise is mixed Gaussian noise.

Filters	ARMSE$_{pos}$ (m)	ARMSE$_{vel}$ (m/s)	Average Iteration Number	Time(ms)
KF	2.1938	0.4118	0	0.0138
HKF	1.4792	0.3784	0	0.0214
MCKF ($\sigma = 2$)	1.4662	0.3777	2.5260	0.0820
STKKF ($v = 3, \sigma = 2$)	1.3965	0.3768	2.7070	0.1094
STKKF ($v = 10, \sigma = 2$)	1.4314	0.3769	2.5898	0.1031
STKKF ($v = 50, \sigma = 2$)	1.4575	0.3775	2.5390	0.1013
MCKF ($\sigma = 3$)	1.6357	0.3838	2.4094	0.0797
STKKF ($v = 3, \sigma = 3$)	1.4837	0.3783	2.5420	0.1011
STKKF ($v = 10, \sigma = 3$)	1.5697	0.3812	2.4602	0.0994
STKKF ($v = 50, \sigma = 3$)	1.6201	0.3832	2.4211	0.0946
MCKF ($\sigma = 5$)	1.9051	0.3963	2.2782	0.0745
STKKF ($v = 3, \sigma = 5$)	1.7110	0.3871	2.3907	0.0980
STKKF ($v = 10, \sigma = 5$)	1.8318	0.3927	2.3236	0.0964
STKKF ($v = 50, \sigma = 5$)	1.8890	0.3955	2.2883	0.0937

6. Discussion

In this paper, the Student's *t* kernel function is employed to replace the traditional Gaussian kernel function in the definition of correntropy to better utilize the heavy-tailed features of noises when the underlying system is disturbed by heavy-tailed non-Gaussian noise. Then the maximum correntropy criterion based on Student's *t* kernel function is applied to the Kalman filter as the optimality criterion. Based on the criterion, a novel Kalman-type filtering algorithm, named STKKF, is derived. Meanwhile, since the fixed-point iteration method is used to update the posterior state estimate, the convergence of the STKKF is also analyzed.

The performance of the proposed filter is verified through comparative simulations with KF, HKF, and MCKF. When both the process noise and measurement noise are Gaussian noises, it can be seen from Table 1 and Figure 1 that the performance of the KF is the best, since both the noises are Gaussian. When the kernel bandwidth is too small, the STKKF may achieve worse performance. However, one can also see that with the increase of the kernel bandwidth, the performance of the STKKF becomes better and approaches that of the KF. This phenomenon can be explained by Theorem 1, that is, when the filter kernel bandwidth $\sigma \to \infty$, the STKKF will tend to become the KF algorithm. In general, with appropriate parameters, the STKKF performance is at least as good as KF. In addition, the average fixed-point iteration numbers required for STKKF to converge and the average implementation times of the filters for one time step are also calculated in Table 1. It is obvious that the average iteration numbers decrease as the kernel bandwidth σ increases, that is, the convergence speed becomes faster. Correspondingly, the average implementation time of STKKF also decreases. In practical real-time applications, the kernel bandwidth should be set appropriately to ensure the algorithm can run in real time.

When the measurement noise is a heavy-tailed (impulsive) non-Gaussian noise, the results shown by Table 2 and Figure 2 demonstrate that with appropriate parameters, the state estimate accuracy of filters based on the maximum correntropy criterion (MCKF and STKKF) outperform that of the other algorithms. However, the implementation times of these maximum correntropy filters are much longer than that of KF and HKF due to their computational complexity. Furthermore, it can be seen that the implementation times of STKKF are consistently longer than MCKF when they have the same kernel bandwidth. This is because the shape of Student's *t* kernel function has a heavier tail than the Gaussian kernel function. In addition, one can also see that when the kernel bandwidth is set as the same, the performance of the STKKF with different *v* is consistently better than MCKF. Additionally, with the increase of the *v*, the performance of the STKKF approaches that of

the MCKF. This phenomenon can be explained by Theorem 2, that is, when $v \to \infty$, the STKKF will tend to become the MCKF with bandwidth σ.

In summary, with proper parameters, the STKKF can outperform the other filtering algorithms, especially for the heavy-tailed non-Gaussian noises. However, like other filters based on the maximum correntropy criterion, the choice of parameters of the kernel function is critical. When the parameters are not appropriate, the filter's performance may degrade, which should be given much attention in practical applications. At present, the Student's t kernel function-based maximum correntropy criterion is only being applied to the linear system. In the future, an extension to the nonlinear system model can be investigated.

Author Contributions: Conceptualization, H.H. and H.Z.; Methodology, H.H.; Software, H.H.; Validation, H.H.; Formal analysis, H.H.; Investigation, H.H. and H.Z.; Resources, H.Z.; Data curation, H.H.; Writing—original draft preparation, H.H.; Writing—review and editing, H.Z.; Visualization, H.H.; Supervision, H.Z.; Project administration, H.Z.; Funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

KF	Kalman filter
EKF	Inertial Measurement Unit
UKF	Unscented Kalman filter
QKF	Quadrature Kalman filter
CKF	Cubature Kalman filter
HKF	Huber-based Kalman filter
MCKF	Maximum correntropy Kalman filter
MEEKF	Minimum error entropy Kalman filter
CKKF	Cauchy kernel-base maximum correntropy Kalman filter
STKKF	Student's t kernel-based maximum correntropy Kalman filter
MCC	Maximum correntropy criterion
UT	Unscented transformation
MSE	Mean square error
PDF	Probability density function

References

1. Duník, J.; Biswas, S.K.; Dempster, A.G.; Pany, T.; Closas, P. State estimation methods in navigation: Overview and application. *IEEE Aerosp. Electron. Syst. Mag.* **2020**, *35*, 16–31. [CrossRef]
2. Huang, Y.; Zhang, Y.; Shi, P.; Wu, Z.; Qian, J.; Chambers, J.A. Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *49*, 2082–2096. [CrossRef]
3. Bizeray, A.M.; Zhao, S.; Duncan, S.R.; Howey, D.A. Lithium-ion battery thermal-electrochemical model-based state estimation using orthogonal collocation and a modified extended Kalman filter. *J. Power Sources* **2015**, *296*, 400–412. [CrossRef]
4. Deng, R.; Xiao, G.; Lu, R.; Liang, H.; Vasilakos, A.V. False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey. *IEEE Trans. Ind. Inform.* **2016**, *13*, 411–423. [CrossRef]
5. Zhao, J.; Mili, L. Robust unscented Kalman filter for power system dynamic state estimation with unknown noise statistics. *IEEE Trans. Smart Grid* **2017**, *10*, 1215–1224. [CrossRef]
6. Sturm, J.; Ennifar, H.; Erhard, S.V.; Rheinfeld, A.; Kosch, S.; Jossen, A. State estimation of lithium-ion cells using a physicochemical model based extended Kalman filter. *Appl. Energy* **2018**, *223*, 103–123. [CrossRef]

7.   Khamseh, H.B.; Ghorbani, S.; Janabi-Sharifi, F. Unscented Kalman filter state estimation for manipulating unmanned aerial vehicles. *Aerosp. Sci. Technol.* **2019**, *92*, 446–463. [CrossRef]

8.   Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*; John Wiley & Sons: Hoboken, NJ, USA, 2004; pp. 381–394.

9.   Julier, S.J.; Uhlmann, J.K. Unscented filtering and nonlinear estimation. *Proc. IEEE* **2004**, *92*, 401–422. [CrossRef]

10. Arasaratnam, I.; Haykin, S.; Elliott, R.J. Discrete-time nonlinear filtering algorithms using Gauss–Hermite quadrature. *Proc. IEEE* **2007**, *95*, 953–977. [CrossRef]

11. Arasaratnam, I.; Haykin, S. Cubature kalman filters. *IEEE Trans. Autom. Control* **2009**, *54*, 1254–1269. [CrossRef]

12. Singh, A.K. Major development under Gaussian filtering since unscented Kalman filter. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 1308–1325. [CrossRef]

13. Huang, Z.; Zhou, N.; Diao, R.; Wang, S.; Elbert, S.; Meng, D.; Lu, S. Capturing real-time power system dynamics: Opportunities and challenges. In Proceedings of the 2015 IEEE Power & Energy Society General Meeting, Denver, CO, USA, 26–30 July 2015; pp. 1–5.

14. Wang, S.; Zhao, J.; Huang, Z.; Diao, R. Assessing Gaussian assumption of PMU measurement error using field data. *IEEE Trans. Power Deliv.* **2017**, *33*, 3233–3236. [CrossRef]

15. Kulikov, G.Y.; Kulikova, M.V. Estimation of maneuvering target in the presence of non-Gaussian noise: A coordinated turn case study. *Signal Process.* **2018**, *145*, 241–257. [CrossRef]

16. Guangcai, W.; Xu, X.; Zhang, T. MM estimation-based robust cubature Kalman filter for INS/GPS integrated navigation system. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–11. [CrossRef]

17. Chang, L.; Hu, B.; Chang, G.; Li, A. Huber-based novel robust unscented Kalman filter. *IET Sci. Meas. Technol.* **2012**, *6*, 502–509. [CrossRef]

18. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy Kalman filter. *Automatica* **2017**, *76*, 70–77. [CrossRef]

19. Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; pp. 123–126.

20. Chen, B.; Zhu, Y.; Hu, J.; Principe, J.C. *System Parameter Identification: Information Criteria and Algorithms*; Newnes: Oxford, UK, 2013; pp. 51–55.

21. Chen, B.; Dang, L.; Gu, Y.; Zheng, N.; Príncipe, J.C. Minimum error entropy Kalman filter. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *6*, 5819–5829. [CrossRef]

22. Erdogmus, D.; Principe, J.C. An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems. *IEEE Trans. Signal Process.* **2002**, *50*, 1780–1786. [CrossRef]

23. Zhang, Y.; Chen, B.; Liu, X.; Yuan, Z.; Principe, J.C. Convergence of a fixed-point minimum error entropy algorithm. *Entropy* **2015**, *17*, 5549–5560. [CrossRef]

24. Liu, X.; Chen, B.; Xu, B.; Wu, Z.; Honeine, P. Maximum correntropy unscented filter. *Int. J. Syst. Sci.* **2017**, *48*, 1607–1615. [CrossRef]

25. Dang, L.; Chen, B.; Wang, S.; Ma, W.; Ren, P. Robust power system state estimation with minimum error entropy unscented Kalman filter. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8797–8808. [CrossRef]

26. Wang, Y.; Zheng, W.; Sun, S.; Li, L. Robust information filter based on maximum correntropy criterion. *J. Guid. Control Dyn.* **2016**, *39*, 1126–1131. [CrossRef]

27. Wang, G.; Li, N.; Zhang, Y. Maximum correntropy unscented Kalman and information filters for non-Gaussian measurement noise. *J. Frankl. Inst.* **2017**, *354*, 8659–8677. [CrossRef]

28. Li, M.; Jing, Z.; Leung, H. Robust Minimum Error Entropy Based Cubature Information Filter With Non-Gaussian Measurement Noise. *IEEE Signal Process. Lett.* **2021**, *28*, 349–353. [CrossRef]

29. Huang, Y.; Zhang, Y.; Li, N.; Chambers, J. Robust Student'st based nonlinear filter and smoother. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 2586–2596. [CrossRef]

30. Cheng, L.; Yue, H.; Xing, Y.; Ren, M. Multipath Estimation Based on Modified ε-Constrained Rank-Based Differential Evolution with Minimum Error Entropy. *IEEE Access* **2018**, *6*, 61569–61582. [CrossRef]

31. Huang, Y.; Zhang, Y.; Zhao, Y.; Shi, P.; Chambers, J.A. A novel outlier-robust Kalman filtering framework based on statistical similarity measure. *IEEE Trans. Autom. Control* **2020**, *66*, 2677–2692. [CrossRef]

32. Wang, J.; Lyu, D.; He, Z.; Zhou, H.; Wang, D. Cauchy kernel-based maximum correntropy Kalman filter. *Int. J. Syst. Sci.* **2020**, *51*, 3523–3538. [CrossRef]

33. Wang, H.; Li, X.; Bi, D.; Xie, X.; Xie, Y. A Robust Student's t-Based Kernel Adaptive Filter. *IEEE Trans. Circuits Syst. II Express Briefs* **2021**, *68*, 3371–3375. [CrossRef]

34. Santamaría, I.; Pokharel, P.P.; Principe, J.C. Generalized correlation function: Definition, properties, and application to blind equalization. *IEEE Trans. Signal Process.* **2006**, *54*, 2187–2197. [CrossRef]

35. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: A localized similarity measure. In Proceedings of the The 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 16–21 July 2006; pp. 4919–4924.

36. Chen, B.; Wang, J.; Zhao, H.; Zheng, N.; Principe, J.C. Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Process. Lett.* **2015**, *22*, 1723–1727. [CrossRef]

37. Agarwal, R.P.; Meehan, M.; O'regan, D. *Fixed Point Theory and Applications*; Cambridge University Press: Cambridge, UK, 2001.
38. Simon, D. Kalman filtering with state constraints: A survey of linear and nonlinear algorithms. *IET Control Theory Appl.* **2010**, *4*, 1303–1318. [CrossRef]