

Article



Maximum Correntropy Unscented Kalman Filter for Spacecraft Relative State Estimation

Xi Liu^{1,*}, Hua Qu^{1,2}, Jihong Zhao¹, Pengcheng Yue¹ and Meng Wang²

- ¹ School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; qh@mail.xjtu.edu.cn (H.Q.); zhaojihong@mail.xjtu.edu.cn (J.Z.); heaven.1452@stu.xjtu.edu.cn (P.Y.)
- ² School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China; avictor.wally@stu.xjtu.edu.cn
- * Correspondence: lx1102@stu.xjtu.edu.cn; Tel.: +86-29-8266-0055

Academic Editor: Vittorio M. N. Passaro

Received: 6 August 2016; Accepted: 8 September 2016; Published: 20 September 2016

Abstract: A new algorithm called maximum correntropy unscented Kalman filter (MCUKF) is proposed and applied to relative state estimation in space communication networks. As is well known, the unscented Kalman filter (UKF) provides an efficient tool to solve the non-linear state estimate problem. However, the UKF usually plays well in Gaussian noises. Its performance may deteriorate substantially in the presence of non-Gaussian noises, especially when the measurements are disturbed by some heavy-tailed impulsive noises. By making use of the maximum correntropy criterion (MCC), the proposed algorithm can enhance the robustness of UKF against impulsive noises. In the MCUKF, the unscented transformation (UT) is applied to obtain a predicted state estimation and covariance matrix, and a nonlinear regression method with the MCC cost is then used to reformulate the measurement information. Finally, the UT is adopted to the measurement equation to obtain the filter state and covariance matrix. Illustrative examples demonstrate the superior performance of the new algorithm.

Keywords: unscented Kalman filter (UKF); unscented transformation (UT); maximum correntropy criterion (MCC)

1. Introduction

The relative state estimation problem is rather important in space communication networks, including inter-satellite link (ISL) establishment, communication support, multiple spacecraft formation flying, and spacecraft rendezvous and docking [1–4]. Many relative motion scenes make use of the radar or/and ladar to obtain the relative measurement information among spacecrafts, which is necessary to determine the relative motions.

As for the state estimation problem, the Kalman filter (KF) is often used for a linear dynamic system, which is, in essence, a recursive minimum l_2 -norm linear filter [5–7]. However, from [8], we know that the relative state estimation in a circular reference orbit involves a linear state Clohessy-Wiltshire (CW) equation and a nonlinear measurement equation. Especially, when the reference orbit is elliptic, the dynamic state equation is also nonlinear [9]. To handle the nonlinear problems, the widely used methods are the extended Kalman filter (EKF) [10–12] and unscented Kalman filter (UKF) [11–13]. The EKF is a popular nonlinear extension of KF, which makes use of the first order Taylor series expansions to approximate the nonlinear system and applies the KF to this approximation. Nevertheless, the approximation is crude and it may lead to filter divergence if the function exhibits highly non-linear characteristics. Furthermore, the derivation of the Jacobian matrices is cumbersome and often results in the implementation difficulties. In UKF, the probability distribution of the state is approximated by a set of deterministically selected sigma points and propagated though

the non-linear process and measurement equations. UKF does not need to calculate the cumbersome Jacobian matrices and provides derivative-free higher-order approximations to achieve more accurate performances than the EKF. Unfortunately, the EKF and UKF may fail to perform well when the measurements are disturbed by some heavy-tailed non-Gaussian noises, which arises naturally in many real applications of engineering. The main reason for this is that both methods are based on the minimum l_2 -norm technique and thus exhibit sensitivity to heavy-tailed noises [14]. To address this problem, some robust methods have been proposed. The Huber's generalized maximum likelihood methodology is very common [15,16], which is a combined minimum l_1 and l_2 -norm technique.

Besides Huber's robust statistics, the optimization criteria in information theoretic learning (ITL) [17,18] provide an alternative effective way, using information theoretic quantities estimated directly from the data as the optimization costs instead of using the usual second-order statistical measures (such as variance and covariance). The ITL costs can capture higher-order statistics of data, and can be used as a robust adaptation cost to achieve excellent performance in a number of applications [19–31]. Particularly, an optimization criterion based on correntropy, called the maximum correntropy criterion (MCC) [17,18], has recently been successfully applied in robust adaptive filtering in the presence of heavy-tailed non-Gaussian noises [17,28–32].

In this paper, a non-linear Kalman type filter based on MCC has been derived, namely the maximum correntropy unscented Kalman filter (MCUKF). In MCUKF, the UT is applied to obtain a predicted state estimation and covariance matrix, and a non-linear regression model under MCC is used to reformulate the measurement information. Then, the UT is adopted to the measurement equation to obtain the filter state and covariance matrix. Similar to the original UKF, the new filter also has a recursive structure and is suitable for online implementation. It is worth noting that the proposed MCUKF is different from the algorithm in [33].

The organization of the paper is as follows. In Section 2, we provide a short review of the MCC and the traditional UKF. In Section 3, the MCUKF algorithm is derived. Two illustrative examples are then provided in Section 4 to show the excellent performance of the MCUKF. Finally, Section 5 gives the conclusion.

2. Preliminaries

2.1. Maximum Correntropy Criterion

Given two random variables $X, Y \in \mathbb{R}$ with joint distribution function $F_{XY}(x, y)$, the correntropy is a generalized similarity measure between *X* and *Y* defined by

$$V(X,Y) = \mathbb{E}\left[\kappa(X,Y)\right] = \int \kappa(x,y) dF_{XY}(x,y),$$
(1)

where E is the expectation operator, and $\kappa(\cdot, \cdot)$ denotes a shift-invariant kernel following the Mercer condition. In this paper, the kernel function is chosen as the Gaussian kernel given by

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

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

In many practical applications, we have only a limited amount of data and the joint distribution F_{XY} is unknown. In these cases, the correntropy can be estimated by the sample mean estimator:

$$\widehat{V}(X,Y) = \frac{1}{N} \sum_{i=1}^{N} G_{\sigma}\left(e(i)\right),\tag{3}$$

where e(i) = x(i) - y(i), with $\{x(i), y(i)\}_{i=1}^{N}$ being *N* samples drawn from F_{XY} .

Taking the Taylor series expansion of the Gaussian kernel yields

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

It can be seen that the correntropy is a weighted sum of all even order moments of the error variable X - Y. In addition, the kernel bandwidth appears as a parameter weighting the second order and higher order moments. It is noted that the correntropy will be dominated by the second order moment when the kernel bandwidth is very large compared to the dynamic range of the data. Given a sequence of error data $\{e(i)\}_{i=1}^{N}$, the cost function of MCC is given by

$$J_{MCC} = \frac{1}{N} \sum_{i=1}^{N} G_{\sigma}\left(e(i)\right).$$
(5)

Assume **W** is a parameter vector of an adaptive system to learn, and let x(i) and y(i) denote the model output and the desired response, respectively. The MCC based learning can be formulated as the following optimization problem:

$$\widehat{\mathbf{W}} = \underset{\mathbf{W}\in\mathbf{\Omega}}{\arg\max} \frac{1}{N} \sum_{i=1}^{N} \mathbf{G}_{\sigma}\left(e(i)\right), \tag{6}$$

where \widehat{W} denotes the optimal solution, and Ω denotes a feasible set of the parameter.

2.2. Unscented Kalman Filter

The UKF provides a suitable alternative for EKF to deal with nonlinear systems. In this paper, we discuss a discrete-time nonlinear system described as

$$\mathbf{x}(k) = \mathbf{f}(k-1, \mathbf{x}(k-1)) + \mathbf{q}(k-1),$$
(7)

$$\mathbf{y}(k) = \mathbf{h}\left(k, \mathbf{x}(k)\right) + \mathbf{r}(k),\tag{8}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ denotes the *n*-dimensional state vector of the system at time step k, $\mathbf{y}(k) \in \mathbb{R}^m$ denotes an *m*-dimensional measurement vector. f represents a nonlinear system function, and h is a nonlinear measurement function and they are assumed to be continuously differentiable. The process noise $\mathbf{q}(k-1)$ and measurement noise $\mathbf{r}(k)$ are mutually uncorrelated, with zero mean values and covariance matrices

$$\mathbf{E}\left[\mathbf{q}(k-1)\mathbf{q}^{T}(k-1)\right] = \mathbf{Q}(k-1), \ \mathbf{E}\left[\mathbf{r}(k)\mathbf{r}^{T}(k)\right] = \mathbf{R}(k).$$
(9)

In general, similar to KF, the UKF includes two steps, namely predict and update:

2.2.1. Predict

First, a set of 2n + 1 sigma points are computed by the estimated state and covariance matrix of last time step:

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

$$\chi^{i}(k-1|k-1) = \widehat{\mathbf{x}}(k-1|k-1) + \left(\sqrt{(n+\lambda)\mathbf{P}(k-1|k-1)}\right)_{i}, i = 1, \dots, n,$$

$$\chi^{i}(k-1|k-1) = \widehat{\mathbf{x}}(k-1|k-1) - \left(\sqrt{(n+\lambda)\mathbf{P}(k-1|k-1)}\right)_{i-n}, i = n+1, \dots, 2n,$$
(10)

where $\left(\sqrt{(n+\lambda)\mathbf{P}(k-1|k-1)}\right)_i$ is the *i*-th column of the matrix square root of $(n+\lambda)\mathbf{P}(k-1|k-1)$, *n* denotes the dimension of the state vector, and λ represents a compound scaling factor defined by

$$\lambda = \alpha^2 (n + \varepsilon) - n, \tag{11}$$

where α determines the extent of the spread of the sigma points, which is usually chosen as $0 < \alpha \le 1$, ε is a scaling factor and is usually set to 3 - n.

The transformed set is given through the process model

$$\chi^{i^*}(k|k-1) = f\left(k-1, \chi^i(k-1|k-1)\right), i = 0, \dots, 2n.$$
(12)

Then, the predicted state mean and associated covariance matrix are calculated by

$$\widehat{\mathbf{x}}(k|k-1) = \sum_{i=0}^{2n} w_m^i \chi^{i^*}(k|k-1),$$
(13)

$$\mathbf{P}(k|k-1) = \sum_{i=0}^{2n} w_c^i \left[\chi^{i^*}(k|k-1) - \widehat{\mathbf{x}}(k|k-1) \right] \\ \times \left[\chi^{i^*}(k|k-1) - \widehat{\mathbf{x}}(k|k-1) \right]^T + \mathbf{Q}(k-1),$$
(14)

where the corresponding weights of the state and covariance matrix are defined as

$$w_m^0 = \frac{\lambda}{(n+\lambda)},$$

$$w_c^0 = \lambda/(n+\lambda) + (1-\alpha^2 + \beta),$$

$$w_m^i = w_c^i = \frac{1}{2(n+\lambda)}, i = 1, \dots, 2n.$$
(15)

with β being related to prior knowledge of the distribution of $\mathbf{x}(k)$ (for Gaussian noises, $\beta = 2$ is optimal).

2.2.2. Update

Similarly, a set of 2n + 1 sigma points are computed by the predicted state mean and covariance matrix

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

$$\chi^{i}(k|k-1) = \widehat{\mathbf{x}}(k|k-1) + \left(\sqrt{(n+\lambda)\mathbf{P}(k|k-1)}\right)_{i}, i = 1, ..., n,$$

$$\chi^{i}(k|k-1) = \widehat{\mathbf{x}}(k|k-1) - \left(\sqrt{(n+\lambda)\mathbf{P}(k|k-1)}\right)_{i-n}, i = n+1, ..., 2n.$$
(16)

The transformed set is given through the measurement model

$$\gamma^{i}(k) = \mathbf{h}(k, \chi^{i}(k|k-1)), i = 0, \dots, 2n.$$
(17)

Then, the predicted measurement mean and covariance matrix are calculated by

$$\widehat{\mathbf{y}}(k) = \sum_{i=0}^{2n} w_m^i \gamma^i(k), \tag{18}$$

$$\mathbf{P}_{\mathbf{y}\mathbf{y}} = \sum_{i=0}^{2n} w_c^i \left[\gamma^i(k) - \widehat{\mathbf{y}}(k) \right] \left[\gamma^i(k) - \widehat{\mathbf{y}}(k) \right]^T + \mathbf{R}(k).$$
(19)

In addition, the state-measurement cross-covariance matrix is computed by

$$\mathbf{P}_{\mathbf{x}\mathbf{y}}(k) = \sum_{i=0}^{2n} w_c^i \left[\chi^i(k|k-1) - \widehat{\mathbf{x}}(k|k-1) \right] \left[\gamma^i(k) - \widehat{\mathbf{y}}(k) \right]^T.$$
(20)

Next, the Kalman gain is given as

$$\mathbf{K}(k) = \mathbf{P}_{\mathbf{x}\mathbf{y}}\mathbf{P}_{\mathbf{y}\mathbf{y}}^{-1}.$$
(21)

Finally, the filter state in this case is obtained by

$$\widehat{\mathbf{x}}(k|k) = \widehat{\mathbf{x}}(k|k-1) + \mathbf{K}(k) \left(\mathbf{y}(k) - \widehat{\mathbf{y}}(k)\right).$$
(22)

Additionally, the filter covariance is recursively updated as

$$\mathbf{P}(k|k) = \mathbf{P}(k|k-1) - \mathbf{K}(k)\mathbf{P}_{\mathbf{y}\mathbf{y}}\mathbf{K}^{T}(k).$$
(23)

3. Unscented Kalman Filter under MCC

To improve the robustness of the UKF, we present the idea to combine the MCC cost function with the UKF framework to derive a novel UKF, which may perform much better in non-Gaussian noise environments, since correntropy contains second and higher order moments of the error.

First, we consider a nonlinear model given by Equation (7) and Equation (8), combine expression Equations (13) and (14) and recast the nonlinear regression model as

$$\begin{bmatrix} \widehat{\mathbf{x}}(k|k-1) \\ \mathbf{y}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{h}(k, \mathbf{x}(k)) \end{bmatrix} + \psi(k),$$
(24)

where $\psi(k)$ denotes by

$$\psi(k) = \begin{bmatrix} \eta(\mathbf{x}(k)) \\ \mathbf{r}(k) \end{bmatrix},$$
(25)

with

$$\Psi(k)$$

$$= \mathbb{E} \left[\psi(k)\psi^{T}(k) \right]$$

$$= \left[\begin{array}{cc} \mathbf{P}(k|k-1) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}(k) \end{array} \right]$$

$$= \left[\begin{array}{cc} \mathbf{T}_{p}(k|k-1)\mathbf{T}_{p}^{T}(k|k-1) & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{r}(k)\mathbf{T}_{r}^{T}(k) \end{array} \right]$$

$$= \mathbf{T}(k)\mathbf{T}^{T}(k),$$
(26)

where $\mathbf{T}(k)$ can be obtained by Cholesky decomposition of $\Psi(k)$. Left multiplying both sides of Equation (24) by $\mathbf{T}^{-1}(k)$, we have the following equation

$$\mathbf{D}(k) = \mathbf{g}(k, \mathbf{x}(k)) + \mathbf{e}(k), \tag{27}$$

5 of 16

where $\mathbf{D}(k) = \mathbf{T}^{-1}(k) \begin{bmatrix} \hat{\mathbf{x}}(k|k-1) \\ \mathbf{y}(k) \end{bmatrix}$, $\mathbf{g}(k, \mathbf{x}(k)) = \mathbf{T}^{-1}(k) \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{h}(k, \mathbf{x}(k)) \end{bmatrix}$, $\mathbf{e}(k) = \mathbf{T}^{-1}(k)\psi(k)$ and the covariance of $\mathbf{e}(k)$ is the identity matrix.

Then, we define the following cost function $J_L(\mathbf{x}(k))$ based MCC:

$$J_L(\mathbf{x}(k)) = \sum_{i=1}^{L} G_{\sigma}(d_i(k) - g_i(k, \mathbf{x}(k))),$$
(28)

where $d_i(k)$ is the *i*-th element of $\mathbf{D}(k)$, $g_i(k, \mathbf{x}(k))$ is the *i*-th row of $g(k, \mathbf{x}(k))$, and L = n + m is the dimension of $\mathbf{D}(k)$.

Under the MCC, the optimal estimate of $\mathbf{x}(k)$ can be found by maximizing the cost function

$$\widehat{\mathbf{x}}(k) = \arg \max_{\mathbf{x}(k)} J_L(\mathbf{x}(k)) = \arg \max_{\mathbf{x}(k)} \sum_{i=1}^L G_\sigma(e_i(k)),$$
(29)

where $e_i(k)$ is the *i*-th element of $\mathbf{e}(k)$ and

$$e_i(k) = d_i(k) - \mathbf{g}_i(k, \mathbf{x}(k)).$$
(30)

Hence, the optimal solution for $\mathbf{x}(k)$ can be found from the following equation

$$\frac{\partial J_L\left(\mathbf{x}(k)\right)}{\partial \mathbf{x}(k)} = 0,\tag{31}$$

which can also be expressed as

$$\sum_{i=1}^{L} \varphi(e_i(k)) \frac{\partial e_i(k)}{\partial x_i(k)} = 0,$$
(32)

when $x_i(k)$ is the *i*-th element of $\mathbf{x}(k)$, and $\varphi(e_i(k)) = G_{\sigma}(e_i(k)) \cdot e_i(k)$. Then, we define $C(k,i) = \varphi(e_i(k))/e_i(k)$ and have

$$C(k,i) = G_{\sigma}(e_i(k)).$$
(33)

Equation (32) further writes in matrix form as

$$\left(\frac{\partial \mathbf{g}(k, \mathbf{x}(k))}{\partial \mathbf{x}(k)}\right)^{T} \mathbf{C}(k) \left[\mathbf{D}(k) - \mathbf{g}(k, \mathbf{x}(k))\right] = 0,$$
(34)

where

$$\mathbf{C}(k) = \begin{bmatrix} \mathbf{C}_{x}(k) & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{y}(k) \end{bmatrix},$$

with

$$\mathbf{C}_{x}(k) = diag\left(\mathbf{G}_{\sigma}\left(e_{1}(k)\right), ..., \mathbf{G}_{\sigma}\left(e_{n}(k)\right)\right), \mathbf{C}_{y}(k) = diag\left(\mathbf{G}_{\sigma}\left(e_{n+1}(k)\right), ..., \mathbf{G}_{n+m}\left(e_{n+m}(k)\right)\right)$$

Based on Equation (27) with only one iteration in [32], we could obtain similar results in the UKF framework by using C(k) to reformulate the measurement information. Similar to [16], there are two ways to do it. One is to re-weight the residual error covariance depending on the value of $e_i(k) = d_i(k) - g_i(k, \mathbf{x}(k))$, the other is to reconstruct the 'pseudo-observation'. In general, the derived results based on two ways are the same. In this paper, we just describe the algorithm based on the first way for simplicity.

Defining $\widetilde{\Psi}(k)$ is the modified covariance, and could be given as

$$\widetilde{\Psi}(k) = \mathbf{T}(k)\mathbf{C}^{-1}(k)\mathbf{T}^{T}(k).$$
(35)

For our next discussion, we write $\widetilde{\Psi}(k)$ into two portions so that

$$\widetilde{\Psi}(k) = \begin{bmatrix} \widetilde{\Psi}_{x}(k) & \mathbf{0} \\ \mathbf{0} & \widetilde{\Psi}_{y}(k) \end{bmatrix}.$$
(36)

In fact, since the true state $\mathbf{x}(k)$ is unknown, we set $\eta(\mathbf{x}(k)) = \hat{\mathbf{x}}(k|k-1) - \mathbf{x}(k) = \mathbf{0}$. In this case, we can see easily that

$$\widetilde{\Psi}_{x}(k) = \mathbf{T}_{p}(k|k-1) * \mathbf{I} * \mathbf{T}_{p}^{T}(k|k-1)$$

$$= \mathbf{P}(k|k-1).$$
(37)

The modified measurement covariance can be obtained

$$\widetilde{\mathbf{R}}(k) = \widetilde{\Psi}_{\boldsymbol{y}}(k). \tag{38}$$

In the above derivation, the MCUKF algorithm can be summarized as follows:

- 1. Choose a proper kernel bandwidth σ ; set an initial estimate $\hat{\mathbf{x}}(0|0)$ and corresponding covariance matrix $\mathbf{P}(0|0)$; and let k = 1;
- 2. Use Equations (10) and (12) to calculate the sigma points and the propagated sigma points with respect to function f;
- 3. Compute predicted estimate $\hat{\mathbf{x}}(k|k-1)$ and covariance matrix $\mathbf{P}(k|k-1)$ by Equations (13) and (14), and adopt Cholesky decomposition to get $\mathbf{T}_p(k|k-1)$;
- 4. Use Equations (16) and (17) to calculate the sigma points and the propagated sigma points with respect to function h;
- 5. Use Equations (24)–(38) to derive the modified $\widetilde{\mathbf{R}}(k)$; compute the predicted measurement mean by Equation (18); and replace $\mathbf{R}(k)$ with $\widetilde{\mathbf{R}}(k)$ in Equation (19) and have the following covariance matrix:

$$\mathbf{P}_{\mathbf{y}\mathbf{y}} = \sum_{i=0}^{2n} w_c^i \left[\gamma^i(k) - \widehat{\mathbf{y}}(k) \right] \left[\gamma^i(k) - \widehat{\mathbf{y}}(k) \right]^T + \widetilde{\mathbf{R}}(k).$$
(39)

6. Use Equations (20)–(23) to compute the filter state mean and covariance matrix; go back to 2.

Remark 1. As one can see, different from the UKF algorithm, the MCUKF uses a non-linear regression model combined MCC to update the measurement information. Note that the kernel bandwidth σ is a key parameter in MCUKF. In general, a smaller kernel bandwidth makes the algorithm more robust (with respect to outliers). However, when the kernel bandwidth is too small, it may result in the filter divergence or accuracy deterioration. The reason for this can be seen in [34]. On the other hand, when σ becomes larger and larger, the MCUKF will behave more and more like the UKF algorithm. In particular, the following theorem holds.

Theorem 1. If the kernel bandwidth σ approaches infinity, the MCUKF will reduce to the traditional UKF.

Proof of Theorem 1. The proof is simple, we give a short explanation. Note that as $\sigma \to \infty$, we have $C(k) \to I$. In this case, the MCUKF is equal to the traditional UKF. \Box

Thus, it is seen that the kernel bandwidth has significant influences on the behavior of MCUKF. In practical applications, the kernel bandwidth can be set manually or optimized by trial and error methods.

4. Illustrative Examples

This section demonstrates the performances of the proposed algorithm by two illustrative examples. In the paper, we mainly compare the estimation performance based on the following benchmarks:

$$MSE_1(m) = \frac{1}{K} \sum_{k=1}^{K} (x(k) - \hat{x}(k|k))^2, m = 1, \dots, M,$$
(40)

MSE =
$$\frac{1}{M} \sum_{m=1}^{M} MSE_1(m)$$
, (41)

$$MSD_{p}(k) = \frac{1}{M} \sum_{m=1}^{M} \left((x(k) - \hat{x}(k|k))^{2} + (y(k) - \hat{y}(k|k))^{2} + (z(k) - \hat{z}(k|k))^{2} \right), k = 1, \dots, K,$$
(42)

$$MSD_{v}(k) = \frac{1}{M} \sum_{m=1}^{M} \left(\left(\dot{x}(k) - \hat{x}(k|k) \right)^{2} + \left(\dot{y}(k) - \hat{y}(k|k) \right)^{2} + \left(\dot{z}(k) - \hat{z}(k|k) \right)^{2} \right), k = 1, \dots, K,$$
(43)

$$TAMSD_{p} = \frac{1}{k_{2} - k_{1}} \sum_{k=k_{1}+1}^{k_{2}} MSD_{p}(k),$$
(44)

TAMSD_v =
$$\frac{1}{k_2 - k_1} \sum_{k=k_1+1}^{k_2} MSD_v(k)$$
, (45)

where *K* is the entire time steps in every Monte Carlo run and *M* represents the total Monte Carlo runs.

4.1. Example 1

Now, a univariate nonstationary growth model (UNGM) that is often used as a benchmark example for nonlinear filtering is considered, whose state and measurement equations are given by

$$x(k) = \alpha_1 x(k-1) + \alpha_2 \frac{x(k-1)}{1 + x(k-1)^2} + \alpha_3 \cos\left(1.2(k-1)\right) + q(k-1),$$
(46)

$$y(k) = \frac{x(k)^2}{20} + r(k).$$
(47)

The parameters are set to $\alpha_1 = 0.5$, $\alpha_2 = 25$, $\alpha_3 = 8$.

First, we consider the case in which all the noises are Gaussian

$$q(k-1) \sim N(0,1),$$

 $r(k) \sim N(0,1).$

In this example, the parameters are chosen as K = 500, M = 100. Table 1 illustrates the MSEs of x, defined in Equation (41), in Gaussian noise. Since all the noises are Gaussian, the UKF gives the smallest MSE in all filters. Moreover, it is noted that when the kernel bandwidth is small, the MCUKF may result in a worse estimation; in contrast, when the kernel bandwidth becomes larger, its performance will approach that of the UKF. Actually, it has been proved easily that when $\sigma \rightarrow \infty$, the MCUKF will reduce to the traditional UKF (see Theorem 1). Therefore, one should choose a larger kernel bandwidth in Gaussian noises.

Second, we change the observation noise into a heavy-tailed non-Gaussian noise, with a mixed-Gaussian distribution

$$q(k-1) \sim N(0,1),$$

 $r(k) \sim 0.8N(0,1) + 0.2N(0,500).$

Table 2 shows the MSEs of *x* in non-Gaussian measurement noise. As one can see, in this case, when kernel bandwidth is too small or too large, the performance of MCUKF will be not good. However, with a proper kernel bandwidth (say $\sigma = 2.0$), the MCUKF can outperform the UKF, achieving the smallest MSE. Again, when σ is very large, MCUKF achieves almost the same performance as the UKF.

Table 1. MSEs of *x* in Gaussian noises.

Filter	MSE of <i>x</i>
UKF	67.6974
MCUKF ($\sigma = 2.0$)	87.6836
MCUKF ($\sigma = 3.0$)	80.8406
MCUKF ($\sigma = 5.0$)	74.0286
MCUKF ($\sigma = 10$)	72.3362
MCUKF ($\sigma = 20$)	68.6795

Table 2. MSEs	of x i	n non-G	aussian	measurement	t noise

Filter	MSE of <i>x</i>
UKF	85.8439
MCUKF ($\sigma = 1.0$)	84.1944
MCUKF ($\sigma = 2.0$)	82.6933
MCUKF ($\sigma = 3.0$)	83.1098
MCUKF ($\sigma = 5.0$)	84.7173
MCUKF ($\sigma = 10$)	85.4411

4.2. Example 2

Finally, we consider a practical example with respect to the relative motion of two spacecrafts, which is illustrated in Figure 1. One of the spacecrafts is called the chief spacecraft, which is moving on the reference orbit, and the other is the deputy spacecraft. They all revolve around the earth and thus the inertial orbital equations of two spacecraft are given as

$$\ddot{\mathbf{r}}_c = -\frac{\mu}{r_c^3} \mathbf{r}_c,\tag{48}$$

$$\ddot{\mathbf{r}}_d = -\frac{\mu}{r_d^3} \mathbf{r}_d,\tag{49}$$

where \mathbf{r}_c and \mathbf{r}_d are the position vectors of the chief spacecraft and the deputy spacecraft in ECI coordinate frame, r_c and r_d are the norms of \mathbf{r}_c and \mathbf{r}_d , respectively, and μ is the gravitational parameter of the earth.



Figure 1. Illustration of example 2.

The position vector of the deputy spacecraft relative to the chief spacecraft is

$$\boldsymbol{a}_{dc} = \boldsymbol{r}_d - \boldsymbol{r}_c. \tag{50}$$

Here, we use the Hill coordinate frame, which is a rectangular, Cartesian, dextral rotating frame centered on the chief spacecraft and refer to it as the local-vertical-local-horizontal (LVLH) frame. Then, the motion of the deputy spacecraft relative to the chief spacecraft in the Hill coordinate frame can be described in [35] as

$$\ddot{x} = 2\omega \dot{y} + \dot{\omega}y + \omega^2 x + \frac{\mu}{r_c^2} - \frac{\mu(r_c + x)}{\left[(r_c + x)^2 + y^2 + z^2\right]^{3/2}} + q_1,$$
(51)

$$\ddot{y} = -2\omega\dot{x} - \dot{\omega}x + \omega^2 y - \frac{\mu y}{\left[(r_c + x)^2 + y^2 + z^2\right]^{3/2}} + q_2,$$
(52)

$$\ddot{z} = -\frac{\mu z}{\left[(r_c + x)^2 + y^2 + z^2\right]^{3/2}} + q_3,$$
(53)

with

$$\ddot{r}_c = r_c \omega^2 - \frac{\mu}{r_c^2},\tag{54}$$

$$\dot{\omega} = -\frac{2\dot{r}_c\omega}{r_c},\tag{55}$$

where x, y and z are the radial, in-track and cross-track coordinates of the Hill frame, and ω denotes the orbital angular velocity of the chief spacecraft.

The radar is set on the chief spacecraft to obtain the measurements, and the measurement coordinate system is showed in Figure 2. The measurement equations are given as

$$\rho = \sqrt{(x^2 + y^2 + z^2)} + r_1, \tag{56}$$

$$\theta = \tan^{-1}\frac{y}{x} + r_2, \tag{57}$$

$$\phi = \tan^{-1} \frac{z}{\sqrt{x^2 + y^2}} + r_3, \tag{58}$$

where ρ is the relative range between the chief spacecraft and deputy spacecraft, θ is the azimuth angle, and ϕ is the elevation angle.



Figure 2. Measurement coordinate system.

The two spacecrafts are on the elliptic low Earth orbits. The initial six orbital elements of the chief spacecraft are showed in Table 3, and the trajectory of which in ECI coordinate frame is propagated by the fixed-step fourth-order Runge–Kutta algorithm. The prediction of filters is performed every 0.1 s and the measurements record from the radar every 1 s. The state vector $\mathbf{x}(k) = \begin{bmatrix} x(k) & y(k) & z(k) & \dot{x}(k) & \dot{y}(k) & \dot{z}(k) \end{bmatrix}^T$ contains the relative position and velocity components in Hill frame, respectively. The initial true state is

$$x(0) = 31.9262$$
 km, $\dot{x}(0) = -0.005583$ km/s,
 $y(0) = -7.1384$ km, $\dot{y}(0) = -0.071774$ km/s,
 $z(0) = 33.4729$ km, $\dot{z}(0) = 0.026249$ km/s,

and the initial estimates and covariance matrix of the states are chosen as

 $\hat{x}(0|0) = 31.9262 \text{ km}, \ \hat{x}(0|0) = -0.004416 \text{ km/s},$ $\hat{y}(0|0) = -8.1384 \text{ km}, \ \hat{y}(0|0) = -0.061774 \text{ km/s},$ $\hat{z}(0|0) = 32.4729 \text{ km}, \ \hat{z}(0|0) = 0.036249 \text{ km/s},$ $\mathbf{P}(0|0) = diag([1,1,1,0.01^2,0.01^2,0.01^2]).$

Table 3. Initial orbital elements of chief spacecraft.

Orbital Elements	Chief Spacecraft
Semi-major axis	8000 km
Eccentricity	0.150
Orbit inclination	$\pi/6$ rad
Argument of perigee	$\pi/6$ rad
Right ascension of the ascending node	$\pi/18$ rad
True anomaly	0 rad

In this example, 100 independent Monte Carlo runs have been conducted, and, in each case, an elapsed time of 7200 s is considered, that is K = 7200, M = 100. Since correntropy is a local similarity measure, the MCUKF may converge to the optimal solution slowly, especially when the initial values deviate greatly from the true values. Thus, we use the UKF during the first 100 s to make the process settle down and then switch to the MCUKF to continue. The performance of EKF, Huber-EKF (HEKF) [36], UKF and novel robust UKF (NRUKF) [16] are shown for comparison with that of the proposed algorithm.

First, we consider the case in which all the noises are Gaussian, that is,

$$\begin{array}{l} q_1 \sim N(0, (10^{-7} \text{ km/s}^2)^2), \ r_1 \sim N(0, (10^{-3} \text{ km})^2), \\ q_2 \sim N(0, (10^{-7} \text{ km/s}^2)^2), \ r_2 \sim N(0, (0.05\pi/180)^2), \\ q_3 \sim N(0, (10^{-7} \text{ km/s}^2)^2), \ r_3 \sim N(0, (0.05\pi/180)^2). \end{array}$$

Figure 3 describes the relative motion of the deputy spacecraft in the Hill coordinate frame, Figures 4 and 5 demonstrate the MSD_p and MSD_v , defined in Equations (42) and (43), with different filters in Gaussian noises, and Table 4 summarizes the corresponding TAMSD_p and TAMSD_v, defined in Equations (44) and (45) (the parameter is set to $k_1 = 1000$, $k_2 = 7200$). Those results illustrate that the UKF play the best performance in all filters in this case and the UKF type filters have the better performance than the EKF type counterparts. One can also observe that the robust filters do not perform as well as their non-robust counterparts in the Gaussian noises. Moreover, it is worth noting that when the kernel bandwidth is small, the MCUKF may achieve a worse performance; whereas, when the kernel bandwidth becomes larger, it has similar results to the UKF.



Figure 3. Relative motion of the deputy spacecraft in the Hill frame.



Figure 4. $\ensuremath{\mathsf{MSD}}_p$ with different filters in Gaussian noises.



Figure 5. MSD_v with different filters in Gaussian noises.

Filter	TAMSD _p	$\mathbf{TAMSD}_{\mathbf{v}}$
EKF	9.6397×10^{-6}	3.7704×10^{-11}
HEKF	$1.0742 imes10^{-5}$	$3.9391 imes 10^{-11}$
UKF	$8.4386 imes10^{-6}$	$2.6552 imes 10^{-11}$
NRUKF	$9.0277 imes 10^{-6}$	$2.7508 imes 10^{-11}$
MCUKF ($\sigma = 2.0$)	$9.3516 imes 10^{-6}$	$2.8038 imes 10^{-11}$
MCUKF ($\sigma = 5.0$)	$8.4909 imes10^{-6}$	$2.6625 imes 10^{-11}$
MCUKF ($\sigma = 20$)	8.4403×10^{-6}	$2.6554 imes 10^{-11}$

Table 4. TAMSD_p and TAMSD_v in Gaussian noises.

Second, we consider the case in which the measurement noises are heavy-tailed, with mixed-Gaussian distributions

$$\begin{aligned} q_1 &\sim N(0, (10^{-7} \text{ km/s}^2)^2), \\ q_2 &\sim N(0, (10^{-7} \text{ km/s}^2)^2), \\ q_3 &\sim N(0, (10^{-7} \text{ km/s}^2)^2), \\ r_1 &\sim 0.9N(0, (10^{-3} \text{ km})^2) + 0.1N(0, (10^{-2} \text{ km})^2), \\ r_2 &\sim 0.9N(0, (0.05\pi/180)^2) + 0.1N(0, (0.5\pi/180)^2), \\ r_3 &\sim 0.9N(0, (0.05\pi/180)^2) + 0.1N(0, (0.5\pi/180)^2). \end{aligned}$$

Figures 6 and 7 reveal the MSD_p and MSD_v with different filters in non-Gaussian noises, and Table 5 lists the corresponding TAMSD_p and TAMSD_v. As one can observe again, the UKF type filters give a smaller TAMSD_p and TAMSD_v than the EKF type counterparts. All the robust filters are superior to their non-robust counterparts in impulsive noises. It is noted that when the kernel bandwidth is very large, MCUKF achieves almost the same performance as the UKF. However, with a proper kernel bandwidth, the MCUKF can outperform the UKF significantly. Particularly, when $\sigma = 2.0$, the MCUKF exhibits the smallest TAMSD_p and TAMSD_v.

Moreover, to compare the computational cost, the computation times of different filters in this example are shown in Table 6. We can see that the computation time of the MCUKF is moderate compared with the UKF, and is superior to that of the HEKF and NRUKF.



Figure 6. MSD_p with different filters in non-Gaussian noises.



Figure 7. MSD_v with different filters in non-Gaussian noises.

Filter	TAMSD _p	TAMSD _v
EKF	1.0680×10^{-4}	$2.2751 imes 10^{-10}$
HEKF	$3.2302 imes10^{-5}$	$7.5898 imes 10^{-11}$
UKF	$4.2884 imes10^{-5}$	$8.4705 imes 10^{-11}$
NRUKF	$2.5510 imes 10^{-5}$	$5.7120 imes 10^{-11}$
MCUKF ($\sigma = 2.0$)	$2.4842 imes 10^{-5}$	$5.6557 imes 10^{-11}$
MCUKF ($\sigma = 5.0$)	$3.1545 imes 10^{-5}$	$6.6535 imes 10^{-11}$
MCUKF ($\sigma = 20$)	$4.1800 imes10^{-5}$	$8.3054 imes 10^{-11}$

Table 5. TAMSD_p and TAMSD_v in non-Gaussian noises.

m 11 /	<u> </u>		•
Table 6.	Computation	time	comparison.
			••••••••••••••••

Filter	Computation Ratio
UKF	1
HEKF	6.81
NRUKF	4.62
MCUKF	3.91

5. Conclusions

In this paper, we propose a new unscented Kalman filter (UKF), namely the maximum correntropy unscented Kalman filter (MCUKF), which shows strong robustness against heavy-tailed non-Gaussian noises. The proposed algorithm is recast in the form of a non-linear regression model and makes use of the MCC to obtain the filter estimates. The MCUKF is then applied to the spacecraft relative state estimation, compared with EKF, HEKF, UKF and NRUKF. Simulation results confirm that, with a large kernel bandwidth, the MCUKF will behave like the original UKF. With a proper kernel bandwidth, the new filter can achieve better performance than other filters, especially when the measurement system is disturbed by some impulsive non-Gaussian noises.

Acknowledgments: This work was supported by the 863 Program (No. 2015AA015702) and the National Natural Science Foundation of China (No. 61371087).

Author Contributions: Xi Liu proposed the main idea, wrote the draft manuscript and presented the experiments; Hua Qu and Jihong Zhao polished the language and were in charge of technical checking; Pengcheng Yue and Meng Wang analyzed the data and drew the figures and tables.

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

References

- Marszalek, M.; Kurz, O.; Drentschew, M.; Schmidt, M.; Schilling, K. Intersatellite Links and Relative Navigation: Pre-conditions for Formation Flights with Pico- and Nanosatellites. In Proceedings of the 18th IFAC World Congress, Milan, Italy, 29 August–3 September 2011; pp. 3027–3032.
- 2. Liu, X.; Sun, Z.; Zhang, J.; Wu, X. Relative Navigation Method of Communication Supporting Spacecraft base-on Particle Filter. *J. Comput. Inf. Syst.* **2013**, *9*, 1–8.
- Wu, F.; Sui, X.; Zhao, Y.; Zhang, Y. Relative navigation for formation flying spacecrafts using X-ray pulsars. In Proceedings of the IEEE/ION Position Location and Navigation Symposium (PLANS), Myrtle Beach, SC, USA, 23–26 April 2012; pp. 1289–1292.
- Kelsey, J.; Byrne, J.; Cosgrove, M.; Seereeram, S.; Mehra, R. Vision-based relative pose estimation for autonomous rendezvous and docking. In Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2006; pp. 1–20.
- 5. Kalman, R.E. A new approach to linear filtering and prediction problems. *Trans. ASME-J. Basic Eng.* **1960**, *82*, 35–45.
- 6. Bryson, A.; Ho, Y. *Applied Optimal Control: Optimization, Estimation and Control;* CRC Press: Boca Raton, FL, USA, 1975.
- 7. Nahi, N.E. Estimation Theory and Applications; Wiley: New York, NY, USA, 1969.
- 8. Hablani, H.B.; Tapper, M.L.; Dana-Bashian, D.J. Guidance and Relative Navigation for Autonomous Rendezvous in a Circular Orbit. *Trans. ASME-J. Basic Eng.* **2002**, *25*, 553–562.
- Baek, K.; Bang, H. Adaptive sparse grid quadrature filter for spacecraft relative navigation. *Acta Astronaut.* 2013, *87*, 96–106.
- 10. Anderson, B.; Moore, J. Optimal Filtering; Prentice-Hall: Englewood Cliffs, NJ, USA, 1979.
- 11. Haykin, S. Kalman Filtering and Neural Networks; John Wiley & Sons: New York, NY, USA, 2001.
- 12. Simon, D. *Optimal State Estimation: Kalman, H*_∞ *and Nonlinear Approaches;* John Wiley & Sons: Hoboken, NJ, USA, 2006.
- 13. Julier, S.; Uhlmann, J.; Durrant-Whyte, H.F. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Autom. Control* **2000**, *45*, 477–482.
- 14. Schick, I.; Mitter, S. Robust Recursive Estimation in the Presence of Heavy-Tailed Observation Noise. *Ann. Stat.* **1994**, *22*, 1045–1080.
- 15. Huber, P. Robust Statistics; Springer: Berlin, Germany, 2011.
- Chang, L.; Hu, B.; Chang, G.; Li, A. Huber-based novel robust unscented Kalman filter. *IET Sci. Meas. Technol.* 2012, 6, 502–509.
- 17. Principe, J.C. Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives; Springer: New York, NY, USA, 2010.
- 18. Chen, B.; Zhu, Y.; Hu, J.; Principe, J.C. *System Parameter Identification: Information Criteria and Algorithms*; Newnes: Oxford, UK, 2013.
- 19. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties, and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298.
- 20. Chen, B.; Principe, J.C. Maximum correntropy estimation is a smoothed MAP estimation. *IEEE Signal Process. Lett.* **2012**, *19*, 491–494.
- Ma, W.; Qu, H.; Zhao, J. Estimator with Forgetting Factor of Correntropy and Recursive Algorithm for Traffic Network Prediction. In Proceedings of the Chinese Control and Decision Conference (CCDC), Guiyang, China, 25–27 May 2013; pp. 490–494.
- 22. Chen, X.; Yang, J.; Liang, J.; Ye, Q. Recursive robust least squares support vector regression based on maximum correntropy criterion. *Neurocomputing* **2012**, *97*, 63–73.
- 23. He, R.; Hu, B.; Yuan, X.; Wang, L. *Robust Recognition via Information Theoretic Learning*; Springer: Amsterdam, The Netherlands, 2014.
- 24. He, R.; Zheng, W.; Hu, B. Maximum correntropy criterion for robust face recognition. *IEEE Trans. Pattern Anal.* **2011**, *33*, 1561–1576.
- 25. He, R.; Hu, B.; Zheng, W.; Kong, X. Robust principal component analysis based on maximum correntropy criterion. *IEEE Trans. Image Process.* **2011**, *20*, 1485–1494.

- 26. Chen, L.; Qu, H.; Zhao, J.; Chen, B.; Principe, J.C. Efficient and robust deep learning with Correntropy-induced loss function. *Neural Comput. Appl.* **2016**, *27*, 1019–1031.
- 27. Bessa, R.J.; Miranda, V.; Gama, J. Entropy, and correntropy against minimum square error in offline, and online three-day ahead wind power forecasting. *IEEE Trans. Power Syst.* **2009**, *24*, 1657–1666.
- 28. Ma, W.; Qu, H.; Gui, G.; Xu, L.; Zhao, J.; Chen, B. Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments. *J. Frankl. Inst.* **2015**, 352, 2708–2727.
- 29. Chen, B.; Xing, L.; Liang, J.; Zheng, N.; Principe, J.C. Steady-State Mean-Square Error Analysis for Adaptive Filtering under the Maximum Correntropy Criterion. *IEEE Signal Process. Lett.* **2014**, *21*, 880–884.
- 30. Shi, L.; Lin, Y. Convex Combination of Adaptive Filters under the Maximum Correntropy Criterion in Impulsive Interference. *IEEE Signal Process. Lett.* **2014**, *21*, 1385–1388.
- 31. Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Principe, J.C. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387.
- 32. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum Correntropy Kalman Filter. Available online: http://arxiv.org/abs/1509.04580 (accessed on 18 September 2016).
- 33. Liu, X.; Chen, B.; Xu, B.; Wu, Z.; Honeine, P. Maximum Correntropy Unscented Filter. Available online: http://arxiv.org/abs/1608.07526 (accessed on 18 September 2016).
- 34. 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.
- 35. Schaub, H.; Junkins, J.L. *Analytical Mechanics of Space Systems*; American Institute of Aeronautics and Astronautics Education Series: Reston, VA, USA, 2003.
- El-Hawary, F.; Jing, Y. Robust Regression-Based EKF for Tracking Underwater Targets. *IEEE J. Ocean. Eng.* 1995, 20, 31–41.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).